

#3

11000 U.S. PTO
09/885315
06/18/01

THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of : Shinichi HAYASHI, et al.

Filed : Concurrently herewith

For : TRAFFIC ENGINEERING METHOD AND....

Serial No. : Concurrently herewith

June 19, 2001

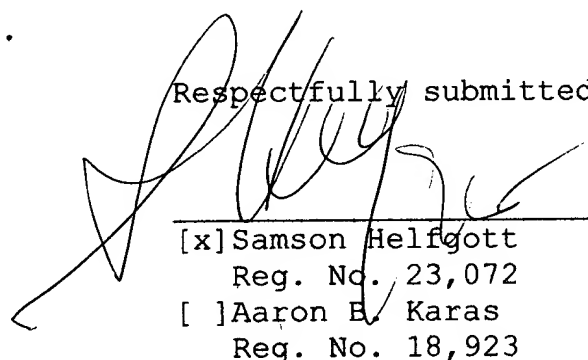
Assistant Commissioner of Patents
Washington, D.C. 20231

SUBMISSION OF PRIORITY DOCUMENT

S I R:

Attached herewith is Japanese Patent Application No.
2000-389077 of December 21, 2000 whose priority has been claimed
in the present application.

Respectfully submitted



☒ Samson Helfgott
Reg. No. 23,072
☐ Aaron E. Karas
Reg. No. 18,923

HELFGOTT & KARAS, P.C.
60th FLOOR
EMPIRE STATE BUILDING
NEW YORK, NY 10118
DOCKET NO.: FUJI 18.659
BHU:priority

Filed Via Express Mail
Rec. No.: EL639693370US
On: June 19, 2001
By: Brendy Lynn Belony
Any fee due as a result of this paper,
not covered by an enclosed check may be
charged on Deposit Acct. No. 08-1634.

日 本 国 特 許 庁
JAPAN PATENT OFFICE



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2000年12月21日

出 願 番 号

Application Number:

特願2000-389077

出 願 人

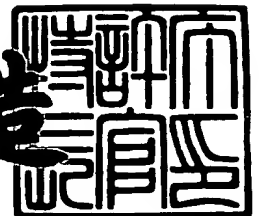
Applicant(s):

富士通株式会社

2001年 4月20日

特 許 庁 長 官
Commissioner,
Japan Patent Office

及 川 耕 造



出証番号 出証特2001-3034308

【書類名】 特許願

【整理番号】 0051753

【提出日】 平成12年12月21日

【あて先】 特許庁長官 及川 耕造 殿

【国際特許分類】 H04L 12/28

【発明の名称】 トラフィックエンジニアリング方法及びそれを用いたノード装置

【請求項の数】 5

【発明者】

【住所又は居所】 福岡県福岡市早良区百道浜2丁目2番1号 富士通西日本コミュニケーション・システムズ株式会社内

【氏名】 林 伸一

【発明者】

【住所又は居所】 福岡県福岡市早良区百道浜2丁目2番1号 富士通西日本コミュニケーション・システムズ株式会社内

【氏名】 木下 博

【特許出願人】

【識別番号】 000005223

【氏名又は名称】 富士通株式会社

【代理人】

【識別番号】 100070150

【住所又は居所】 東京都渋谷区恵比寿4丁目20番3号 恵比寿ガーデンプレイスタワー32階

【弁理士】

【氏名又は名称】 伊東 忠彦

【電話番号】 03-5424-2511

【手数料の表示】

【予納台帳番号】 002989

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9704678

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 トラフィックエンジニアリング方法及びそれを用いたノード装置

【特許請求の範囲】

【請求項 1】 ネットワーク全体が複数ノードの集合体であるエリアに分割されており、前記ネットワークのトラフィックエンジニアリングを行う方法において、

各エリア毎にエリア内で閉じたロードバランシングを実行することを特徴とするトラフィックエンジニアリング方法。

【請求項 2】 全体が複数ノードの集合体であるエリアに分割され、かつ、全体の資源の最適化がトラフィックエンジニアリングにより行われるネットワークを構成するノード装置において、

各エリア毎にエリア内で閉じたロードバランシングを行うためのエリア内宛先を決定するエリア内宛先決定手段を有することを特徴とするノード装置。

【請求項 3】 請求項 2 記載のノード装置において、

外部からパケットが供給される入り口ノードを構成するノード装置は、外部から供給されたパケットのアドレス情報を基にロードバランシングを行うための正規化値を演算し、前記正規化値を前記パケットのスイッチング情報に付加するスイッチング情報生成手段とを有することを特徴とするノード装置。

【請求項 4】 請求項 2 記載のノード装置において、

エリアの境界にあるエリア境界ノードを構成するノード装置は、隣接エリアから供給されるパケットのスイッチング情報から、自エリア内で閉じたロードバランシングを実行するための正規化値を抽出する正規化値抽出手段を有することを特徴とするノード装置。

【請求項 5】 全体が複数ノードの集合体であるエリアに分割され、かつ、全体の資源の最適化がトラフィックエンジニアリングにより行われるネットワークを構成するノード装置において、

障害を検出したときロードバランシングを実行している最も近い上流側のノードに障害を通知する障害通知手段を有することを特徴とするノード装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、トラフィックエンジニアリング方法及びそれを用いたノード装置に関し、ネットワーク内におけるトラフィックエンジニアリング方法及びそれを用いたノード装置に関する。

【0002】

現在のインターネット上では、従来のデータ通信だけでなく音声や映像のようになりリアルタイムサービスに至るまで、実に多種の情報がやり取りされるようになって来た。その結果、インターネット上のトラフィックは年々急速に増加しており、輻輳問題の解決が不可欠となっている。

【0003】

【従来の技術】

複数のノードで構成されるネットワークにおいて、パケットをソースからディスティネーションにフォワーディングする最適ルートを自律的に決定するルーティングプロトコルにはRIP (Routing Information Protocol)、OSPF (Open Shortest Path First)、BGP4 (Border Gateway Protocol Version 4)、IS-IS (Intermediate System to Intermediate System) などが存在する。今日のネットワークでは、これらのプロトコルを使用して最適ルートを求め、このルート上でパケットフォワーディングを行っている。

【0004】

ノードがパケットをフォワーディングする際、一般的には図1に示すように、各ノード10、11、12がパケットのディスティネーションアドレスを参照して行う。

【0005】

これより高速なパケットフォワーディングを行うための技術としてMPLS (Multi Protocol Label Switching) に代表され

るカットスルー方式が注目されている。MPLSでは、図2に示すように、ルーティングプロトコルにより算出された最適ルート上にLSP (Label Switched Path) が設定される。ここで、LSPの両端に位置するノード15, 17をエッジノードと呼び、そのエッジノード間 (MPLSドメイン内) に位置するLSP上の各ノード16をコアノードと呼ぶ。

【0006】

次に、LDP (Label Distribution Protocol) を用いて、LSP上の各ノードが転送先方路を決めるためのラベルが配布される。こうして、MPLSドメインの外部より転送されて来たパケットに対し、まず発側のエッジノードがラベルを付加してからLSP上にパケット載せて転送を行い、最終的に着側エッジノードでラベルが取り除かれてMPLSドメインの外部に転送されるまでは、途中のコアノードではラベルのみを参照したレイヤ2での転送を行うだけで良いため、高速な転送処理が可能となる。

【0007】

このようにルーティングプロトコルとMPLS技術を使用することにより、最適なルートを使用した高速フォワーディングが実現できるのであるが、今日のインターネットのように加入者増により爆発的にトラフィックが増加すると、輻輳やパケットロスが発生してしまう。MPLSは、このように高速フォワーディングが可能というメリットがあるものの、トラフィックが集中した場合にIPルーティングのようなソフトウェアによる臨機応変な経路制御ができないため、輻輳やパケットロスが発生するというデメリットもある。この輻輳やパケットロスの発生を防止するために、ネットワーク全体の資源の最適化を自動的に行う制御であるトラフィックエンジニアリング (TE) が利用出来る。

【0008】

トラフィックエンジニアリング機能自体は特定のレイヤ2媒体に依存しないのであるが、上記MPLSのように発側および着側ノード間でLSPを設定するようなネットワーク上での利用が最も効果的である。トラフィックエンジニアリングの負荷分散方式については、例えば特願平12-12195号に記載のものがある。この方式は、図3に示すように、発側ノード20から着側エッジノード2

1に複数のマルチパスLSP1, LSP2, LSP3を設定し、このマルチパスでトラフィックの分配を行うことで、ネットワーク全体のトラフィックの平均化を行っている。

【0009】

この技術では、トラフィックの現在の負荷を認識するため、図4に示すように、各ノードはリンク毎の平均使用率を算出し、全ノードに対して定期的に広告（Flooding）を行う。発側ノード20は、広告により受信した各ノードのリンク毎の平均使用率をもとに、LSP毎の実際のトラフィック（実効ロード）を算出し、全LSPの実効ロードが同一値に近づくように、マイクロフロー毎にトラフィックの移動を行い、平均化を実現している。マイクロフローとは、あるエンドユーザ間で使用されるフローを指し、これに対し共通の宛先を持つマイクロフローをまとめたものはアグリゲートフローと呼ぶ。

【0010】

マイクロフローをどのLSPにマッピングするかを選択は、図5（A）に示すLSP決定テーブルを用いて実施され、新規マルチパスの追加の度に、このLSP決定テーブル内で分割される領域数も増加していく。まず、発側ノード20にてパケット内のアドレス情報をキーとして正規化値を算出する。その正規化値にてLSP決定テーブルをインデックスし、マッピングするLSPを決定する。このとき、トラフィックの平均化を行うためには、図5（A）に示す状態から図5（B）に示すように、LSP決定テーブルのLSPの境界を移動することでマッピングするLSPの切り替えを行うことにより、トラフィック分散を実現している。

【発明が解決しようとする課題】

従来のトラフィックエンジニアリング技術では、発側ノードが全ノードから定期的に送られてくる各リンク毎の平均使用率を収集し、それを基にLSP毎の実効ロードを算出した上で一括してトラフィックの分散処理を行っていた。したがって、小規模ネットワークでのロードバランシングは可能であったが、OSPFルーティングプロトコルなどのように複数エリアを収容する大規模ネットワークになると、発側ノードへの負担が重過ぎるためロードバランシングの適用は不可

能であった。

【 0 0 1 1 】

また、複数ルートのうち何れかのルートが障害になった場合、従来のトラフィックエンジニアリングでは、発側ノードがネットワークトポロジの変化やLDPのリフレッシュ機能でしか検出できないため、その間は障害ルートを含めた複数パスで負荷分散を実施し続けることになり、トラフィックの高速な救済ができずTCPコネクションなどのマイクロフローなどは救済することが不可能であった。

【 0 0 1 2 】

本発明は、上記の点に鑑みなされたものであり、大規模ネットワークであっても高速にロードバランシングを実行し、障害が発生した場合も障害ルートのトラフィックロス的高速に救済することができるトラフィックエンジニアリング方法及びそれを用いたルータ装置を提供することを目的とする。

【 0 0 1 3 】

【課題を解決するための手段】

請求項1に記載の発明は、ネットワーク全体が複数ノードの集合体であるエリアに分割されており、前記ネットワークのトラフィックエンジニアリングを行う方法において、

各エリア毎にエリア内で閉じたロードバランシングを実行するため、

大規模ネットワークであってもロードバランシング実行ノードが必要とするメモリ容量を大幅に削減でき高速にロードバランシングを実行することができる。

【 0 0 1 4 】

請求項2に記載の発明は、全体が複数ノードの集合体であるエリアに分割され、かつ、全体の資源の最適化がトラフィックエンジニアリングにより行われるネットワークを構成するノード装置において、

各エリア毎にエリア内で閉じたロードバランシングを行うためのエリア内宛先を決定するエリア内宛先決定手段を有するため、

エリア内で閉じたロードバランシングを実行することが可能となり、大規模ネットワークであってもロードバランシング実行ノードが必要とするメモリ容量を

大幅に削減でき高速にロードバランシングを実行することができる。

【0015】

請求項3に記載の発明は、請求項2記載のノード装置において、

外部からパケットが供給される入り口ノードを構成するノード装置は、外部から供給されたパケットのアドレス情報を基にロードバランシングを行うための正規化値を演算し、前記正規化値を前記パケットのスイッチング情報に付加するスイッチング情報生成手段とを有するため、

エリア境界ノードに正規化値を通知することが可能となる。

【0016】

請求項4に記載の発明は、請求項2記載のノード装置において、

エリアの境界にあるエリア境界ノードを構成するノード装置は、隣接エリアから供給されるパケットのスイッチング情報から、自エリア内で閉じたロードバランシングを実行するための正規化値を抽出する正規化値抽出手段を有するため、

エリア境界ノードでは抽出した正規化値を用いてロードバランシングを行うことができ、プロトコルの識別やヘッダチェックなどを行う必要がなく高速のロードバランシングを行うことができる。

【0017】

請求項5に記載の発明は、全体が複数ノードの集合体であるエリアに分割され、かつ、全体の資源の最適化がトラフィックエンジニアリングにより行われるネットワークを構成するノード装置において、

障害を検出したときロードバランシングを実行している最も近い上流側のノードに障害を通知する障害通知手段を有するため、

ロードバランシングを実行しているノードでトラフィックの分配を高速に行うことが可能となる。

【0018】

付記6に記載の発明は、請求項2記載のノード装置において、

障害通知を受信した前記入り口ノードまたはエリア境界ノードは、障害が発生したルートに流していたトラフィックを他のルートに分散しなおすトラフィック分散手段を有するため、

障害ルートを流れていたトラフィックでのトラフィックロスを高速に救済することが可能となる。

【 0 0 1 9 】

付記 7 に記載の発明は、付記 6 記載のノード装置において、

障害通知を受信した前記入り口ノードまたはエリア境界ノードは、障害が発生したルートに流していたトラフィックを他のルートに分散しなおしたときトラフィックロスが発生するか否かを判定する障害通知受信手段を有するため、

障害が発生したルートに流していたトラフィックを他のルートに分散可能か否かを認識できる。

【 0 0 2 0 】

付記 8 に記載の発明は、付記 7 記載のノード装置において、

前記トラフィック分散手段は、前記障害通知受信手段でトラフィックロスが発生すると判定したとき、前記障害が発生したルートに流していたトラフィックを新たに設定したルートに切り替えるため、

障害ルート以外のルートが高トラフィックの場合でも、障害ルートに流していたトラフィックを救済することができる。

【発明の実施の形態】

図 6 は、本発明のトラフィックエンジニアリング方法を適用したネットワークシステムの一実施例の構成図を示す。

【 0 0 2 1 】

このネットワークではルーティングプロトコルとして O S P F を使用する。O S P F では、ネットワーク全体が複数ノードの集合体であるエリアに分割される。図 6 においては、ネットワークはエリア 2 5, 2 6, 2 7 から構成されており、エリア 2 5 はノード 2 5 a ~ 2 5 e より構成され、エリア 2 6 はノード 2 5 d, 2 5 e, 2 6 a ~ 2 6 c より構成され、エリア 2 7 はノード 2 6 c, 2 7 a ~ 2 7 c より構成されている。ノード 2 5 a が入り口ノード (I n g r e s s ノード)、ノード 2 7 c が宛先ノード (E g r e s s ノード) とすると、入り口エリア 2 5 と宛先エリア 2 7 の間は、バックボーンエリア 2 6 で接続される 2 階層のトポロジー構造をとる。ここで、各エリアの境界に位置するノード 2 5 d、2 5

e, 26cそれぞれは、OSPFの機能により自ノードをエリア境界ノード（ABR）として認識する。また、このネットワーク内では、高速なスイッチングを行うためのカットスルー方式としてMPLSを使用する。

【0022】

図7は、本発明方法に適用される発側エッジノードの一実施例のブロック構成図を示す。なお、本実施例では図6の入り口ノード25aが対応する。

【0023】

同図中、L3インタフェース部31は、外部ネットワークから送られたIPパケット等のレイヤ3の受け付け処理を行う。バッファ部32は、各ノードが受信したレイヤ3のパケットについて、スイッチング情報を付加して次ノードに向けて送信するまでの間、パケット情報を保持しておく。L2インタフェース部33は、レイヤ3ヘッダ情報等のマイクロフローを識別する情報が反映されたレイヤ3パケットを指定出力ポートより次ノードに向けて送信する。

【0024】

正規化値演算部34は、パケット情報から、実トラフィックの特性（ソースアドレスやデステネーションアドレス）をもとに、ロードバランシングを実行するための正規化値を算出する。トラフィック分散部35は、算出された正規化値に従って、トラフィックをTEバスグループ内のどのルートへ割当てていくかを決定する。また、ルート上の障害発生時に、当該障害ルート以外の全ルートによるロードバランシングを実行することでトラフィックロスが発生するか否かの判定結果を受け、トラフィックロスが発生しないと認識した場合は、当該障害ルートに流れていたトラフィックをそれ以外の複数ルートに分散しなおす。逆に、トラフィックロスが発生すると認識した場合は、新たにルートを設定し当該障害ルートに流れていたトラフィックを、新たに設定したルートに切り替えてロードバランシングを実行する。

【0025】

出力ポート／スイッチング情報決定部36は、各ノードにおいて、指定された宛先およびパスに対応するパケット出力ポートを決定し、スイッチング情報の生成に必要な各パラメータを決定する。ルーティング制御部37は、デスティネー

ションアドレスから自ノードのルーティング情報を検索し、どの宛先ノードへ向けたルートでロードバランシングを実行するかを選択する。エリア内宛先決定部 38 は、ロードバランシングを実行する際に必要な、ネットワーク全体における宛先ではなく、エリア内の閉じた範囲での宛先決定を行う。

【0026】

スイッチング情報生成部 39 は、次ノードに向けたスイッチング情報を生成し、その情報にソースアドレスやデスティネーションアドレスを基に算出した正規化計算値を付加する。障害通知受信部 40 は、障害検出ノードから通知される障害発生通知を受信し、障害ルート以外の全ルートによるロードバランシングを実行することでトラフィックロスが発生するか否かの判定を行う。トラフィック管理部 41 は、エリア内の全ノードから広告されたトラフィック情報を保持する。パス設定／解放部 43 は、TE パスグループを構成するパスの設定および削除処理を行う。

【0027】

宛先アドレスルックアップテーブル 44 は、宛先アドレスから宛先ノード情報を求めるための検索テーブルである。エリア内宛先決定テーブル 45 は、宛先ノード情報からエリア内のどの宛先までの範囲でロードバランシングを行うかを決定するための検索テーブルである。閾値テーブル 46 は、エリア内でロードバランシングを行うために、トラフィックの分散先（TE パスグループおよび LSP）を求めるための検索テーブルである。スイッチング情報決定テーブル 47 は、エリア内の宛先およびトラフィックの分散先に関する情報をもとに、パケットの出力先とスイッチング情報を求めるための検索テーブルである。

【0028】

本発明において新規に追加された部位はエリア内宛先決定部 38、障害通知受信部 40 であり、必要データとしてエリア内宛先決定テーブル 45 も追加されている。また、機能が追加された部位はトラフィック分散部 35、スイッチング情報生成部 39 である。

【0029】

入り口側エッジノードであるノード 25a での処理について説明する。L3 イ

インタフェース部 31 では、企業や ISP などの外部ネットワークからのレイヤ 3 パケットを受付け、パケット内のアドレス情報を正規化値演算部 34 と L3 ルーティング制御部 37 にそれぞれ通知した後、パケット情報をバッファ部 32 に保持させる。

【0030】

アドレス情報を受けた L3 ルーティング制御部 37 では、宛先アドレスをもとに宛先アドレスルックアップテーブル 44 を検索してパスの終端である宛先ノードを特定し、その結果をエリア内宛先決定部 38 に通知する。従来は宛先ノードまで範囲でロードバランシングを実施していたのに対し、本発明では各エリアに閉じた範囲でロードバランシングを行うため、エリア内宛先決定部 38 では、L3 ルーティング制御部 37 で求めた宛先ノード情報をもとにエリア内宛先決定テーブル 45 を検索してエリア内での宛先（エリア境界ノード）を判定し、その結果を出力ポート／スイッチング情報決定部 36 とトラフィック分散部 35 に通知する。

【0031】

一方、L3 インタフェース部 31 よりアドレス情報を受けた正規化値演算部 34 では、それらの情報を正規化関数にかけ、ロードバランシングを実行するための正規化値を算出し、その値をトラフィック分散部 35 とスイッチング情報生成部 39 に通知する。

【0032】

次に、トラフィック分散部 35 では、エリア内宛先決定部 38 で求めたエリア内の宛先と、正規化値演算部 34 で算出した正規化値をもとに閾値テーブル 46 を参照して、トラフィックを TE パスグループ内のどのルートへ割当てするかを決定し、その結果を出力ポート／スイッチング情報決定部 36 に通知する。また、トラフィック管理部 41 は、自エリア内の全ノードから定期的に広告されるトラフィック情報を L2 インタフェース部 33 より受け取って管理し、トラフィック分散部 35 にトラフィック情報を通知する。

【0033】

トラフィック分散部 35 およびエリア内宛先決定部 38 から通知された情報を

もとに、出力ポート／スイッチング情報決定部 3 6 ではスイッチング情報決定テーブル 4 7 を検索し、パケットを次ノードに転送するための出力ポートと、パケット内に設定するスイッチング情報の決定を行った上でスイッチング情報生成部 3 9 に通知する。

【 0 0 3 4 】

ここで、従来は、出力ポート／スイッチング情報決定部からの通知情報をもとにスイッチング情報生成部で次ノードへのスイッチング情報を生成し、L 2 インタフェース部に渡していたが、本発明では、通常のスイッチング情報に正規化値演算部 3 4 で算出した正規化値も付加した上で L 2 インタフェース部 3 3 に通知する。その後、スイッチング情報を受けた L 2 インタフェース部 3 3 では、その内容をバッファ部 3 2 で保持していたパケットに反映させた後、出力ポートより次ノードに向けて送信を行う。

【 0 0 3 5 】

図 8 は、本発明方法に適用されるコアノードの一実施例のブロック構成図を示す。なお、本実施例では図 6 におけるエリア 2 5 のノード 2 5 b, 2 5 c, 2 5 d, 2 5 e、エリア 2 6 のノード 2 6 a, 2 6 b, 2 6 c、エリア 2 7 のノード 2 7 a, 2 7 b, 2 7 c が対応する。

【 0 0 3 6 】

同図中、L 2 インタフェース部（上流側）5 1 は、トラフィックの上流側のノードから送信されたパケットの受け付け処理を行う。また、自ノードがルート上の障害を検出した場合は、ロードバランシングを実行する最も近いノードに対して、障害通知を送信する。バッファ部 5 2 は、各ノードが受信したパケットを、スイッチング情報を編集して次ノードに向けて送信するまでの間、保持しておく。L 2 インタフェース部（下流側）5 3 は、トラフィックの下流側のスイッチング情報が反映されたパケットを指定出力ポートより次ノードに向け送信する。

【 0 0 3 7 】

正規化値抽出部 5 4 は、スイッチング情報内に設定されたロードバランシングを実行するための正規化値を抽出する。トラフィック分散部 5 5 は、抽出された正規化値に従って、トラフィックを T E パスグループ内のどのルートへ割当て

かを決定する。また、ルート上の障害発生時に、当該障害ルート以外の全ルートによるロードバランシングを実行することでトラフィックロスが発生するか否かの判定結果を受け、トラフィックロスが発生しないと認識した場合は、当該障害ルートに流していたトラフィックをそれ以外の複数ルートに分散しなおす。逆に、トラフィックロスが発生すると認識した場合は、新たにルートを設定し、当該障害ルートに流していたトラフィックを新たに設定したルートに切り替えてロードバランシングを実行する。

【 0 0 3 8 】

出力ポート／スイッチング情報決定部 5 6 は、各ノードにおいて、指定された宛先およびパスに対応するパケット出力ポートを決定し、スイッチング情報の生成に必要な各パラメータを設定する。L 2 ルーティング制御部 5 7 は、デスティネーションアドレスから自ノードのルーティング情報を検索し、どの宛先ノードへ向けたルートでロードバランシングを実行するかを選択する。エリア内宛先決定部 5 8 は、ロードバランシングを実行する際に必要な、ネットワーク全体における宛先ではなく、エリア内の閉じた範囲での宛先決定を行う。

【 0 0 3 9 】

スイッチング情報生成部 5 9 は、次ノードに向けたスイッチング情報を生成し、その情報にソースアドレスやデスティネーションアドレスを基に算出した正規化計算値を付加する。障害通知受信部 6 0 は、障害検出ノードから通知される障害発生通知を受信し、障害ルート以外の全ルートによるロードバランシングを実行することでトラフィックロスが発生するか否かの判定を行う。トラフィック管理部 6 1 は、エリア内の全ノードから広告されたトラフィック情報を保持する。障害通知部 6 2 は、あるルート上で何らかの障害が発生した場合、当該障害を検出したノードよりロードバランシングを実行している最も近い上流ノードに向けて障害の発生を通知する。パス設定／解放部 6 3 は、T E パスグループを構成するパスの設定および削除処理を行う。

【 0 0 4 0 】

宛先アドレスルックアップテーブル 6 4 は、宛先アドレスから宛先ノード情報を求めるための検索テーブルである。エリア内宛先決定テーブル 6 5 は、宛先ノ

ード情報からエリア内のどの宛先までの範囲でロードバランシングを行うかを決定するための検索テーブルである。閾値テーブル 6 6 は、エリア内でロードバランシングを行うために、トラフィックの分散先（TE パスグループおよび LSP）を求めるための検索テーブルである。スイッチング情報決定テーブル 6 7 は、エリア内の宛先およびトラフィックの分散先に関する情報をもとに、パケットの出力先とスイッチング情報を求めるための検索テーブルである。

【 0 0 4 1 】

本発明において新規に追加された部位は、正規化抽出部 5 4、エリア内宛先決定部 5 8、障害通知受信部 6 0、障害通知部 6 2 であり、必要データとしてエリア内宛先決定テーブル 6 5 も追加されている。また、機能が追加された部位は、トラフィック分散部 5 5 である。なお、コアノードとエッジノードに分けて説明を行ったが、同一機器で異なる設定をすることでも各ノードの機能を実現可能である。

【 0 0 4 2 】

コアノードでの処理について説明する。L2 インタフェース部 5 1 では、前ノードから転送されたパケットを受付け、パケット内のスイッチング情報を正規化値抽出部 5 4 および L2 ルーティング制御部 5 7 にそれぞれ通知した後、パケット情報をバッファ部 5 2 に保持させる。

【 0 0 4 3 】

自ノードがロードバランシングを実行するように設定されたノードで、しかもパス上のエリア境界ノードであった場合、L2 ルーティング制御部 5 7 では L2 インタフェース部 5 1 より通知されたスイッチング情報をもとに宛先アドレスルックアップテーブル 6 4 を検索してパスの終端である宛先ノードを特定し、その結果をエリア内宛先決定部 5 8 に通知する。エリア内宛先決定部 5 8 では、エッジノードと同様に、L2 ルーティング制御部 5 7 で求めた宛先ノード情報をもとにエリア内宛先決定テーブル 6 5 を検索してエリア内での宛先（エリア境界ノードまたは宛先ノード）を判定し、その結果を出力ポート／スイッチング情報決定部 5 6 とトラフィック分散部 5 5 に通知する。

【 0 0 4 4 】

従来は入り口ノードのみがロードバランシング用の正規化値を入手する必要があったのに対し、本発明ではエリア内で閉じたロードバランシングを実施するため、正規化値抽出部 54 が、L2 インタフェース部 51 より受け取ったパケット内情報より、自エリア内に閉じてロードバランシングを実行するために必要な正規化値を抽出し、その結果をトラフィック分散部 55 に通知する。

【0045】

次に、トラフィック分散部 55 では、エリア内宛先決定部 58 で求めたエリア内の宛先と、正規化値抽出部 54 で抽出した正規化値をもとに閾値テーブル 66 を参照して、トラフィックをTEパスグループ内のどのルートへ割当てするかを決定し、その結果を出力ポート／スイッチング情報決定部 56 に通知する。またトラフィック管理部 61 は、自エリア内の全ノードから定期的に広告されるトラフィック情報をL2 インタフェース部 53 より受け取って管理し、トラフィック分散部 55 にトラフィック情報を通知する。

【0046】

これに対し、自ノードがロードバランシング実行ノードでない場合は、上記における正規化値抽出部 54、トラフィック分散部 55 の各処理は不要である。以降の出力ポート／スイッチング情報決定部 56、スイッチング情報生成部 59、L2 インタフェース部 53 の各処理はエッジノードと同様である。また、宛先ノード 27c については、ロードバランシングとは直接関係しないため説明は省略する。ところで、ロードバランシング実行ノードとは、コアノードの内でロードバランシングを実行するエリア境界ノードのことである。なお、エリア境界ノードであっても、LSP 上に位置しないノードであればロードバランシングは実行しない。

【0047】

これにより、エリア内で閉じたロードバランシングの実施が可能になり、大規模なネットワークにおいても最適なトラフィックエンジニアリングを行うことができる。また、エッジノードにおいて算出した正規化値をパケットに入れてエリア境界ノードに通知することで、カットスルーのパケット転送方式の利点を生かしたまま、高速フォアワーディングにてロードバランシングを実行できる。

【0048】

次に、上流側エッジノードでの障害処理について説明する。障害通知受信部40では、あるルート上で障害が発生した場合に、障害を検出したコアノードより送信される障害通知をL2インタフェース部33より受取る。そして、当該障害ルートに流していたトラフィックを、当該障害ルート以外の複数ルートに分散しなおした場合にトラフィックロスが発生するか否かを、広告されてきた各ノードのトラフィック情報をもとに判定し、その判定結果をトラフィック分散部35に通知する。

【0049】

通知を受けたトラフィック分散部35では、トラフィックロスが発生しないという判定結果であった場合、当該障害ルートのトラフィックをそれ以外の全ルートに分配しなおす。逆に、トラフィックロスが発生するという判定結果を受けた場合は、新たに別ルートを追加設定した上で、当該障害ルートに流していたトラフィックを、新ルートに切り替えてロードバランシングを行う。

【0050】

次に、コアノードでの障害処理について説明する。障害通知部62では、ロードバランシング処理を行っているルート上での障害発生についてL2インタフェース部53より通知を受け、ロードバランシングを実行している最も近い上流ノード、つまり入り口ノードあるいはパス上のエリア境界ノードに対して、障害通知をL2インタフェース部51から送出する。

【0051】

一方、自ノードがロードバランシング実行ノードであり、障害検出ノードより上記の障害通知を受けた場合の処理は、エッジノードと同様の処理を行う。これにより、ルート上で障害が発生した場合でも、当該障害を検出したノードが、ロードバランシングを実行しているノードに障害を通知して、当該障害以外の複数ルートでトラフィックを分散しなおすことで、高速なトラフィックロスの救済が可能となる。

【0052】

また、当該障害以外のルートへトラフィックを分散しなおすとトラフィックロ

スが発生すると判断した場合、新たにルートを設定し、新ルートに当該障害ルートのトラフィックを切り替えることで、障害ルートのトラフィック救済が可能となる。

【 0 0 5 3 】

更に具体的な実施例について説明する。

【 0 0 5 4 】

まず、OSPFで算出した最適ルートに従ってLSP（デフォルトパス）を設定する。LSP設定用のプロトコルには、LDPやRSVP-LSP-Tunnel（RSVPのMPLS拡張版）等いくつかの方法が提案されているが、RSVP-LSP-Tunnelを使用するものとし、その場合のデフォルトパス設定処理の一実施例のフローチャートを図9に示す。

【 0 0 5 5 】

図9において、パス設定要求を認識したパス設定／解放部43, 63は、ステップS10でOSPFにてエッジノード間の最適ルートを算出する。ステップS12で、図10に示すように入り口ノード25aから宛先ノード27cに向けて、算出された最適ルートに沿ったPathメッセージをL2インタフェース部33, 53より送信する。その応答として宛先ノード27cから逆ルートで受信メッセージ（Resvメッセージ）が送信され、入り口ノード25aが最終的にこの受信メッセージを受け取ることで、パケットフォワーディング用のデフォルトパスが設定される。

【 0 0 5 6 】

このデフォルトパスに沿ってトラフィックが流れ始めると、LSP上の各ノード25a, 25d, 26a, 26c, 27a, 27cでは、自ノードにおけるトラフィックの現在の負荷状況を認識するために、図11に示すように、ハードウェアにより出側ポートの物理リンク（物理チャネル）毎のパケット転送数およびパケット廃棄数を統計情報として周期的に収集し、それらをもとに物理リンク毎の使用率を算出して統計情報に加える。

【 0 0 5 7 】

各ノードは、上記で求めた物理リンク毎の統計情報（平均使用率を含む）を、

OSPFのOpaque LSAを使用して、自エリア内の他ノード全てに対して広告する。なお、Opaque LSAとは、OSPFのメッセージによって各ノード間の状態を交換する際にメッセージのパケットに含まれるLSA（リンク状態広告）を使用者が目的に応じて汎用的に使えるよう拡張したものである。図12に、OSPFのOpaque LSAのフォーマットを示す。OSPFのOpaque LSAでは、OSPFパケットヘッダ、Opaque LSAの後に、カード毎の統計情報がカード数だけ格納される。

【0058】

図13に広告の様子を示す。同図中、エリア#10内のノードAは全ての隣接ノードB、C、DにOpaque LSAを広告し、ノードB、C、Dも受信したOpaque LSAを全ての隣接ノードに広告するが、受信済みのノードは同じOpaque LSAを受信したときは破棄する。

【0059】

この広告情報を収集して管理するのが、従来は入り口ノードのみであったのに対し、本発明では入り口ノードだけでなく、LSP上で自分をエリア境界ノードとして認識しているノードもロードバランシングを行う必要があるため、それらのノードのトラフィック管理部41、61も広告情報を収集して管理する。

【0060】

次に、ロードバランシング実行ノードの処理について説明する。図14は、ロードバランシング処理の第1実施例のフローチャートを示す。同図中、ステップS20でロードバランシング実行ノードであるか否かを判別し、ロードバランシング実行ノードである場合にのみステップS22に進む。ステップS22でロードバランシング実行ノードのトラフィック分散部35、55は、広告により受信したエリア内各ノードのリンク毎の平均使用率をトラフィック管理部41、61より受取り、それをもとにTE（トラフィックエンジニアリング）パスグループのトラフィックを算出する。

【0061】

ここで言うTEパスグループとは、デフォルトパスとそれに対する全TEマルチパスを合わせて一つのグループとしたものを意味し、従来から用いられている

考え方である。TEパスグループの算出例としては、図15に示すように、あるLSP1（デフォルトパス）が複数リンク（Link 1、2、…i…n）で構成されている場合、広告で得られた各リンクのトラフィック情報をもとにLSP1の実効負荷を算出し、さらにLSP1を含むTEパスグループ（LSP1、2、…i…n）としての使用率を算出する。

【0062】

ステップS24でトラフィック分散部41、61は、図16に示すようにTEパスグループの使用率を定期的に算出し、その実効負荷がある上限の閾値を一定時間連続して越えていたらデフォルトパスが輻輳していると判断する。輻輳と判断すると、ステップS26でパス設定／解放部43、63に指示を送り、TEマルチパスを追加設定する。

【0063】

ここまでの処理で従来と異なるのは、従来のTEマルチパスの設定範囲がデフォルトパスと同様に入り口ノードから宛先ノードまでであったのに対し、本発明では図17に示すように、各エリア25、26、27それぞれの閉じた範囲でTEマルチパスの設定を行う点である。なお、図17中、実線はデフォルトパスのLSPを示し、一点鎖線は既存のマルチパスのLSPを示し、破線は新規追加のマルチパスのLSPを示しており、ノード25a、25d、26cがロードバランシング実行ノードである。

【0064】

こうして、ロードバランシング実行ノードでは、輻輳状況に応じて上記の処理を繰り返し、トラフィックがさらに増加してTEパスグループが輻輳したと判断する度に、新たなTEマルチパスを追加し、逆に一定時間トラフィックが下限の閾値を下回れば輻輳が解除されたと判断して、ステップS28でTEマルチパスを削除する。

【0065】

次に、図18は、ロードバランシング処理の第2実施例のフローチャートを示す。この処理は、入り口ノード25aがMPLSドメイン外よりL3インタフェース部31を介してパケットを受信した場合に実行される。

【 0 0 6 6 】

同図中、ステップ S 3 0 で自ノードが入り口ノードか否かを判別し、入り口ノードの場合はステップ S 3 2 に進み、L 3 ルーティング制御部 3 7 がパケットから抽出した宛先アドレス（デスティネーションアドレス）をもとに、図 1 9 に示す構成の宛先アドレスルックアップテーブル 4 4 を検索して、パスの終端である宛先ノードに対応した識別子（Associate Pointer）を決定する。そして、得られた識別子をもとにエリア内宛先決定部 3 8 が図 2 0 に示す構造のエリア内宛先決定テーブル 4 5 を検索し、ロードバランシングの宛先に対応した識別子（L. B. Table Pointer）を決定する。ここで言う宛先とは、従来の宛先ノードまでの宛先ではなく、エリア内でロードバランシングを実施するための宛先である。その後、ステップ S 3 4 で受信パケット内の IP ソースアドレス及び IP デスティネーションアドレスをもとに、正規化値演算部 3 4 がハッシュ関数（CRC 1 6）を用いて正規化値（0 ～ 6 5 5 3 5）を求めステップ S 4 2 に進む。

【 0 0 6 7 】

一方、入り口ノードでない場合はステップ S 3 6 に進み、自ノードがロードバランシング実行ノードであるか否かを判別する。ロードバランシング実行ノードの場合、つまり、エリア境界ノードである場合にはステップ S 3 8 に進み、MP L S がパケットに付加するラベルの値をもとに L 2 ルーティング制御部 5 7 が宛先アドレスルックアップテーブルを検索して同様の識別子を得る。なお、検索には CAM などの特殊メモリを持つハードウェアを用いて高速化しても良い。そして、得られた識別子をもとにエリア内宛先決定部 5 8 がエリア内宛先決定テーブル 6 5 を検索し、ロードバランシングの宛先に対応した識別子（L. B. Table Pointer）を決定する。その後、ステップ S 4 0 でスイッチング情報に付加されている正規化値を抽出してステップ S 4 2 に進む。

【 0 0 6 8 】

ステップ S 4 2 では、上記のエリア内宛先に対応する識別子をもとに、トラフィック分散部 3 5, 5 5 が図 2 1 に示す構造の閾値テーブル 4 6, 6 6 を検索した先の領域には、該当 T E パスグループ内の各 L S P に対して、どの割合でロー

ドバランシングを実施するべきかを示す閾値が設定されている。例えばデフォルトパス 1 本、T E マルチパス 2 本で構成される T E パスグループがロードバランシングを行っていた場合、図 2 3 (A) に示すように、閾値テーブル 4 6, 6 6 を検索した先の領域は 2 つの負荷分散境界値によって L S P 1, L S P 2, L S P 3 に対応した 3 領域に分割される。

【 0 0 6 9 】

ここで、上記で求めた L S P 1, L S P 2, L S P 3 の総領域は、正規化値 (0 ~ 6 5 5 3 5) の範囲で割当てられているため、トラフィック分散部 3 3, 3 5 が正規化値と閾値テーブル 4 6, 6 6 内の領域を照らし合わせることで、そのトラフィックが T E パスグループ内のどの L S P に負荷分散されるかが一意に決まる。また、図 1 4 のフローチャートにおいて T E マルチパスが追加または削除される度に、この領域の負荷分散境界の数も増減するため、それに従い領域の再設定が行われる。

【 0 0 7 0 】

ステップ S 4 4 では、出力ポート／スイッチング情報決定部 3 6, 5 6 は、エリア内宛先決定部 3 8, 5 8 よりエリア内のロードバランシング実行宛先を受け取り、トラフィック分散部 3 5, 5 5 よりどの L S P がトラフィックの分散先であるかの情報を受け取り、それをもとに図 2 2 に示す構造のスイッチング情報決定テーブル 4 7, 6 7 を検索して、パケットを次ノードに出力する際に付加するラベル情報およびパケットの出力ポートを求める。

【 0 0 7 1 】

次ノードに送出するパケットのスイッチング情報はスイッチング情報生成部 3 9, 5 9 で生成するが、ステップ S 4 5 で自ノードが入り口ノードか否かを判別し、入り口ノードの場合にのみステップ S 4 6 に進み、図 2 3 (B) に示すように、通常のスイッチング情報の他に、ハッシュ演算により算出した正規化値も一緒に付加する。これが従来と異なる点である。

【 0 0 7 2 】

これは、本発明ではエリア境界ノードもロードバランシングを実行するため、パケットを受信したエリア境界ノードの正規化抽出部 5 4 が、図 2 3 (C) に示

すように、入り口ノードがパケットに付加した正規化値を参照することで、エリア境界ノードも入り口ノードと同様の付加分散処理が行えるようにするものである。つまり、ロードバランシングノードがエリア境界ノードの場合は、新たに宛先アドレスから正規化値を算出する必要がなく、入り口ノードにて計算された値を流用するだけで済む為、より高速なフォワーディングが実現できる。

【 0 0 7 3 】

上記の手法により複数パスによるロードバランシングが実行されている場合を前提として、パス上で障害が発生した状況について説明する。

【 0 0 7 4 】

図 2 4 に示すように、ロードバランシングを実行中のマルチパス上でリンクまたはノード障害が発生した場合、従来はデフォルト 3 0 秒の比較的長い周期により隣接ノード間で送出し合う O S P F の H e l l o プロトコル等を利用してトポロジー変化を認識し障害を検出していたのに対し、本発明では、既存のハード機能であるリンク毎の L O S (L o s s o f s i g n a l) あるいは L O F (L o s s o f f r a m e) の検出をトリガに障害通知を行うことで、より早い段階で障害を認識する。図 2 4 においては、コアノード 2 6 a が障害を認識している。

【 0 0 7 5 】

図 2 5 は、障害を検出したノードの障害通知部 6 2 が実行する障害検出処理の一実施例のフローチャートを示す。同図中、ステップ S 5 0 で自ノードがロードバランシング実行ノードか否かを判別し、ロードバランシング実行ノードの場合は通知を行う必要が無いため、後述の障害通知受信処理を実行する。バランシング実行ノードでなければステップ S 5 4 で、ロードバランシングを実施している最も近い上流側ノード（入り口ノードまたはエリア境界ノード）宛に障害の発生を通知する。図 2 6 では障害を認識したコアノード 2 6 a がエリア境界ノード 2 5 d に障害の発生を通知している。

【 0 0 7 6 】

障害の通知方法としては、図 2 7 に示す構造の R S V P - L S P - T u n n e l の R e s v T e a r メッセージを L 2 インタフェース部 5 1 から送信するこ

とで可能である。その他にもホップ by ホップによるリンク単位での通知や、トラフィックと逆方向に障害通知用の L S P を設定する方法も考えられる。R e s v T e a r メッセージを使用する場合の例として、パス設定時に送信する P a t h メッセージの S E S S I O N オブジェクト内にロードバランシング実行ノードのアドレスを設定する。このロードバランシング実行ノードのアドレスは、図 2 8 に示すように P a t h メッセージがロードバランシング実行ノードを通過する度に順次置き換えられる。パス上の各ノードは、P a t h メッセージ通過時にこのアドレスを自ノードに保存しておき、障害検出時には、保存されたロードバランシング実行ノードのアドレスに向けて R e s v T e a r メッセージを送信することで障害を通知できる。

【 0 0 7 7 】

図 2 9 は、障害通知を受信したロードバランシング実行ノードが実行する障害通知受信処理の一実施例のフローチャートを示す。

【 0 0 7 8 】

同図中、ステップ S 6 0 で自ノードがロードバランシング実行ノードか否かを判別し、ロードバランシング実行ノードであれば、ステップ S 6 2 で障害通知受信部 6 0 で当該障害ルートに流していたトラフィックについて、当該障害ルート以外の全ルートでの負荷分散が実行可能かどうか判定を行う。この判定方法としては、ロードバランシング時に収集した各 L S P の使用率を利用する。

【 0 0 7 9 】

例えば図 3 0 (A) に示すように T E パスグループ内の各 L S P 1 (1 0 M b p s 中の 6 M b p s を使用) 、 L S P 2 (3 0 M b p s 中の 2 5 M b p s を使用) 、 L S P 3 (1 0 M b p s 中の 4 M b p s を使用) において、図に示す使用率でロードバランシングを実行中に、L S P 1 上で障害が発生したとする。この場合、T E パスグループ全体での空き帯域 (1 5 M b p s) から L S P 1 の空き帯域 (4 M b p s) を引いた結果と、L S P 1 の実行負荷 (6 M b p s) とを比較する。この場合、引いた結果が当該障害ルート (L S P 1) の実効負荷よりも大いので、再分散を実施してもトラフィックロスが発生しないものとステップ S 6 2 で判定し、ステップ S 6 4 でトラフィック分散部 3 5 により当該障害ルート以

外の全ルート L S P 2, L S P 3 に対してトラフィックが再分散される。この時、閾値テーブル 4 6 については、図 3 1 (A) に示すように負荷分散境界の数を 1 つ減らし (当該障害 L S P 1 の領域が削除されるため)、残りの L S P 2, L S P 3 で領域の再分配を行う。この様子を図 3 2 (A) に模式的に示す。

【 0 0 8 0 】

逆に、ステップ S 6 2 で図 3 0 (B) に示すように、再分散すればトラフィックロスが発生すると判定した場合、ステップ S 6 6 でトラフィック分散部 3 5 よりパス設定／解放部 4 3 に対して、新たにエリア内に閉じた T E マルチパスを追加設定するよう指示が出される。新しい T E マルチパスの追加方法は、通常のロードバランシングにおける T E マルチパスの追加処理と同様である。T E マルチパスを追加した後、ステップ S 6 8 で当該障害ルートに流していたトラフィックが新たに設定した L S P のルートに対してトラフィック分散部 3 5 により切替えられる。この時の閾値テーブルの領域は図 3 1 (B) に示すように変更される。この様子を図 3 2 (B) に模式的に示す。

【 0 0 8 1 】

以上の処理を通して、ロードバランシングを実施しているパス上で障害が起きた際のトラフィックロスを高速に救済することが可能になる。つまり、例えば O S P F のネットワークを介したユーザ間で T C P を使用したサービス (T e l n e t 等) 実施中にルート上で障害が発生した場合、通常であれば障害を検出復旧するまで A c k を受信出来ないため、コネクションが切断されてしまう可能性があるが、高速な障害検出トラフィック救済を実施することで、それが回避可能となる。

【 0 0 8 2 】

このように、O S P F のエリアなどの階層化の概念を用いた大規模なルーティングプロトコルネットワークにて、エリア内で閉じたロードバランシングを行うことにより、各ノードはエリア内の全トラフィックデータを持つだけでよく、トラフィックエンジニアリングに使用する全エリアのデータを保持する必要がないため、ロードバランシング実行ノードが必要とするメモリ容量を大幅に削減でき、既存技術では不可能であった大規模ネットワークにおいても最適なトラフィッ

クエンジニアリングが可能となる。

【0083】

また、エッジノードにてソースアドレスやデスティネーションアドレスなどを基にロードバランシングを行う正規化値を算出し、当該正規化値をエリア境界ノードに通知し、エリア境界ノードは当該正規化値を使用して、ロードバランシングを実行することにより、エリア境界ノードにてIPやIPXなどのプロトコルの識別やIPプロトコルなどのヘッダチェックなどを行う必要がないため、カットスルー方式の packets 転送方式の利点を生かしたまま高速フォワーディングにてロードバランシングが可能となる。

【0084】

また、複数のルートを用いてロードバランシングを行っている状態において、1つのルートで何らかの障害が発生した場合、当該障害を検出したノードが、ロードバランシングを実行しているノードに障害を通知し、障害通知を受信したロードバランシング実行ノードが当該障害ルートに流していたトラフィックを、それ以外の複数ルートに分配することにより、障害ルートを流れていたトラフィックでのトラフィックロスを高速に救済することが可能となる。また、これまでは一変化を待って、ユーザトリガの接続再確立が必要であったが、トラフィックロスを高速に救済できることから、アグリゲートされたTCP接続などのマイクロフローの救済も可能となる。

【0085】

更に、障害通知を受信したロードバランシング実行ノードが当該障害ルートに流していたトラフィックを、それ以外の複数ルートに分配するとトラフィックロスが発生すると判断した場合、新たにルートを設定し、当該障害ルートに流していたトラフィックを、新たに設定したルートに切替を行うことにより、障害ルート以外のルートが高トラフィックの場合でも、障害ルートに流していたトラフィックを救済することができ、より信頼性の高い接続レス packets 転送サービスを提供可能となる。

【0086】

なお、エリア宛先決定部38、58が請求項記載のエリア宛先決定手段に

対応し、スイッチング情報生成手段がスイッチング情報生成部 3 9 に対応し、正規化値抽出部 5 4 が正規化値抽出手段に対応し、障害通知部 6 2 が障害通知手段に対応し、障害通知受信部 4 0 が障害通知受信手段に対応する。

【 0 0 8 7 】

(付記 1) ネットワーク全体が複数ノードの集合体であるエリアに分割されており、前記ネットワークのトラフィックエンジニアリングを行う方法において、

各エリア毎にエリア内で閉じたロードバランシングを実行することを特徴とするトラフィックエンジニアリング方法。

【 0 0 8 8 】

(付記 2) 全体が複数ノードの集合体であるエリアに分割され、かつ、全体の資源の最適化がトラフィックエンジニアリングにより行われるネットワークを構成するノード装置において、

各エリア毎にエリア内で閉じたロードバランシングを行うためのエリア内宛先を決定するエリア内宛先決定手段を有することを特徴とするノード装置。

【 0 0 8 9 】

(付記 3) 付記 2 記載のノード装置において、

外部からパケットが供給される入り口ノードを構成するノード装置は、外部から供給されたパケットのアドレス情報を基にロードバランシングを行うための正規化値を演算し、前記正規化値を前記パケットのスイッチング情報に付加するスイッチング情報生成手段とを有することを特徴とするノード装置。

【 0 0 9 0 】

(付記 4) 付記 2 記載のノード装置において、

エリアの境界にあるエリア境界ノードを構成するノード装置は、隣接エリアから供給されるパケットのスイッチング情報から、自エリア内で閉じたロードバランシングを実行するための正規化値を抽出する正規化値抽出手段を有することを特徴とするノード装置。

【 0 0 9 1 】

(付記 5) 全体が複数ノードの集合体であるエリアに分割され、かつ、全

体の資源の最適化がトラフィックエンジニアリングにより行われるネットワークを構成するノード装置において、

障害を検出したときロードバランシングを実行している最も近い上流側のノードに障害を通知する障害通知手段を有することを特徴とするノード装置。

【 0 0 9 2 】

(付記 6) 付記 2 記載のノード装置において、

障害通知を受信した前記入り口ノードまたはエリア境界ノードは、障害が発生したルートに流していたトラフィックを他のルートに分散しなおすトラフィック分散手段を有することを特徴とするノード装置。

【 0 0 9 3 】

(付記 7) 付記 6 記載のノード装置において、

障害通知を受信した前記入り口ノードまたはエリア境界ノードは、障害が発生したルートに流していたトラフィックを他のルートに分散しなおしたときトラフィックロスが発生するか否かを判定する障害通知受信手段を有することを特徴とするノード装置。

【 0 0 9 4 】

(付記 8) 付記 7 記載のノード装置において、

前記トラフィック分散手段は、前記障害通知受信手段でトラフィックロスが発生すると判定したとき、前記障害が発生したルートに流していたトラフィックを新たに設定したルートに切り替えることを特徴とするノード装置。

【 0 0 9 5 】

【発明の効果】

上述の如く、請求項 1 に記載の発明は、各エリア毎にエリア内で閉じたロードバランシングを実行するため、大規模ネットワークであってもロードバランシング実行ノードが必要とするメモリ容量を大幅に削減でき高速にロードバランシングを実行することができる。

【 0 0 9 6 】

請求項 2 に記載の発明は、各エリア毎にエリア内で閉じたロードバランシングを行うためのエリア内宛先を決定するエリア内宛先決定手段を有するため、エリ

ア内で閉じたロードバランシングを実行することが可能となり、大規模ネットワークであってもロードバランシング実行ノードが必要とするメモリ容量を大幅に削減でき高速にロードバランシングを実行することができる。

【 0 0 9 7 】

請求項 3 に記載の発明は、外部からパケットが供給される入り口ノードを構成するノード装置は、外部から供給されたパケットのアドレス情報を基にロードバランシングを行うための正規化値を演算する正規化値演算手段と、正規化値を、パケットのスイッチング情報に付加するスイッチング情報生成手段とを有するため、エリア境界ノードに正規化値を通知することが可能となる。

【 0 0 9 8 】

請求項 4 に記載の発明は、エリアの境界にあるエリア境界ノードを構成するノード装置は、隣接エリアから供給されるパケットのスイッチング情報から、自エリア内で閉じたロードバランシングを実行するための正規化値を抽出する正規化値抽出手段を有するため、エリア境界ノードでは抽出した正規化値を用いてロードバランシングを行うことができ、プロトコルの識別やヘッダチェックなどを行う必要がなく高速のロードバランシングを行うことができる。

【 0 0 9 9 】

請求項 5 に記載の発明は、障害を検出したときロードバランシングを実行している最も近い上流側のノードに障害を通知する障害通知手段を有するため、ロードバランシングを実行しているノードでトラフィックの分配を高速に行うことが可能となる。

【 0 1 0 0 】

付記 6 に記載の発明は、障害通知を受信した入り口ノードまたはエリア境界ノードは、障害が発生したルートに流していたトラフィックを他のルートに分散しなおすトラフィック分散手段を有するため、障害ルートを流れていたトラフィックでのトラフィックロスが高速に救済することが可能となる。

【 0 1 0 1 】

付記 7 に記載の発明は、障害通知を受信した入り口ノードまたはエリア境界ノードは、障害が発生したルートに流していたトラフィックを他のルートに分散し

なおしたときトラフィックロスが発生するか否かを判定する障害通知受信手段を有するため、障害が発生したルートに流していたトラフィックを他のルートに分散可能か否かを認識できる。

【 0 1 0 2 】

付記 8 に記載の発明は、トラフィック分散手段は、障害通知受信手段でトラフィックロスが発生すると判定したとき、障害が発生したルートに流していたトラフィックを新たに設定したルートに切り替えるため、障害ルート以外のルートが高トラフィックの場合でも、障害ルートに流していたトラフィックを救済することができる。

【図面の簡単な説明】

【図 1】

従来の一般的な I P パケットフォワーディング処理を示す図である。

【図 2】

従来 of MPL S による I P パケットフォワーディング処理を示す図である。

【図 3】

従来 of トラフィックエンジニアリング方式を説明するための図である。

【図 4】

従来 of トラフィック分散方式を説明するための図である。

【図 5】

L S P 決定テーブル of L S P の境界を移動を説明するための図である。

【図 6】

本発明 of トラフィックエンジニアリング方法を適用したネットワークシステムの一実施例の構成図である。

【図 7】

本発明方法に適用される発側エッジノードの一実施例のブロック構成図である。

【図 8】

本発明方法に適用されるコアノードの一実施例のブロック構成図である。

【図 9】

デフォルトパス設定処理の一実施例のフローチャートである。

【図 1 0】

デフォルトパス設定を説明するための図である。

【図 1 1】

ハードウェアによるトラフィック情報収集を示す図である。

【図 1 2】

OSPFのOpaqueLSAのフォーマットを示す図である。

【図 1 3】

広告の様子を示す図である。

【図 1 4】

ロードバランシング処理の第 1 実施例のフローチャートである。

【図 1 5】

TEパスグループの使用率算出を説明するための図である。

【図 1 6】

輻輳状況の監視と判定を説明するための図である。

【図 1 7】

エリア単位のTEマルチパスの設定を示す図である。

【図 1 8】

ロードバランシング処理の第 2 実施例のフローチャートである。

【図 1 9】

宛先アドレスルックアップテーブル 4 4 の構成を示す図である。

【図 2 0】

エリア内宛先決定テーブル 4 5 の構造を示す図である。

【図 2 1】

構造の閾値テーブル 4 6, 6 6 の構造を示す図である。

【図 2 2】

スイッチング情報決定テーブル 4 7, 6 7 の構造を示す図である。

【図 2 3】

付加分散実施を説明するための図である。

【図 2 4】

ロードバランシング中での障害検出を説明するための図である。

【図 2 5】

障害を検出したノードの障害通知部 6 2 が実行する障害検出処理の一実施例のフローチャートである。

【図 2 6】

ロードバランシング実行ノードへの障害通知を説明するための図である。

【図 2 7】

RSVP-LSP-Tunnel の Resv Tear メッセージの構造を示す図である。

【図 2 8】

各ノードへの障害通知を説明するための図である。

【図 2 9】

障害通知を受信したロードバランシング実行ノードが実行する障害通知受信処理の一実施例のフローチャートである。

【図 3 0】

再分散によるトラフィックロス発生判定を説明するための図である。

【図 3 1】

再分散による閾値テーブルの変化を説明するための図である。

【図 3 2】

障害検出後の負荷分散を説明するための図である。

【符号の説明】

3 1 L 3 インタフェース部

3 2, 5 2 バッファ部

3 3, 5 1, 5 3 L 2 インタフェース部

3 4 正規化値演算部

3 5, 5 5 トラフィック分散部

3 6, 5 6 出力ポート／スイッチング情報決定部

3 7, 5 7 ルーティング制御部

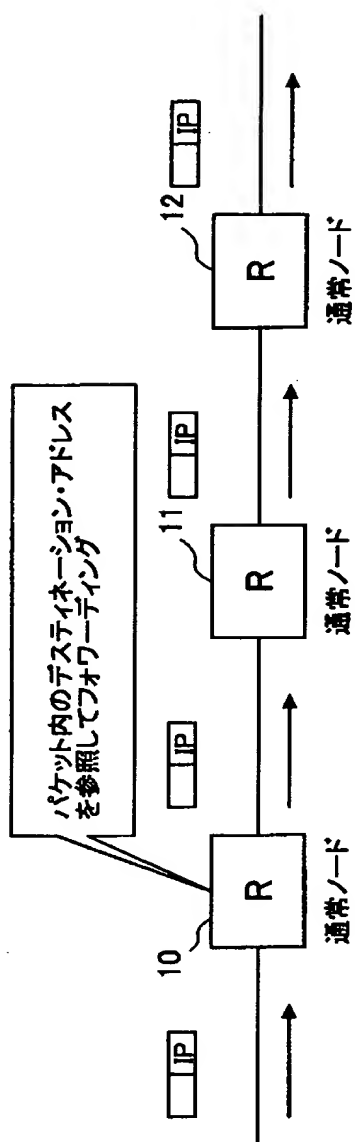
- 3 8, 5 8 エリア内宛先決定部
- 3 9, 5 9 スイッチング情報生成部
- 4 0, 6 0 障害通知受信部
- 4 1, 6 1 トラフィック管理部
- 4 3, 6 3 パス設定／解放部
- 4 4, 6 4 宛先アドレスルックアップテーブル
- 4 5, 6 5 エリア内宛先決定テーブル
- 4 6, 6 6 閾値テーブル
- 4 7, 6 7 スイッチング情報決定テーブル
- 5 4 正規化値抽出部
- 6 2 障害通知部

【書類名】

図面

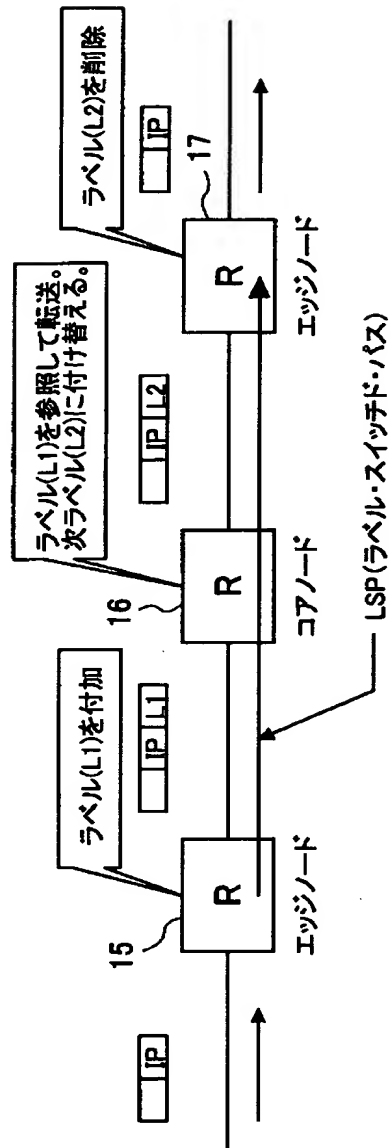
【図 1】

従来の一般的な IP パケットフォワーディング処理を示す図



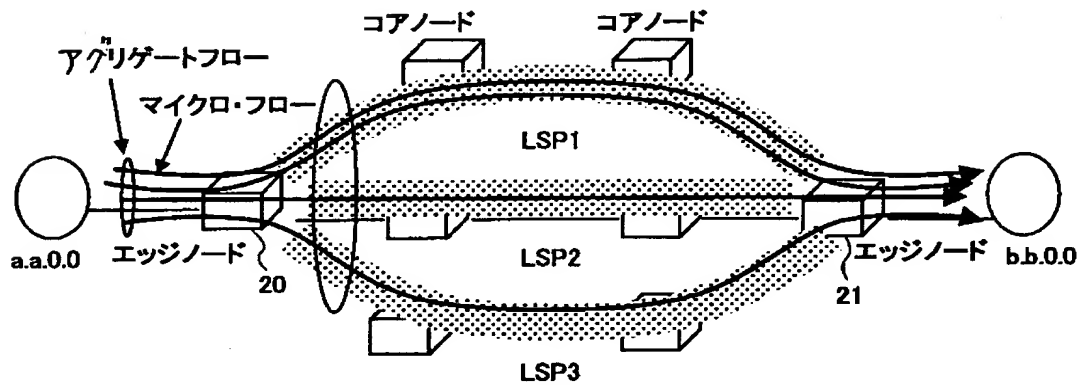
【図 2】

従来のMPLSによるIPパケットフォワーディング処理を示す図



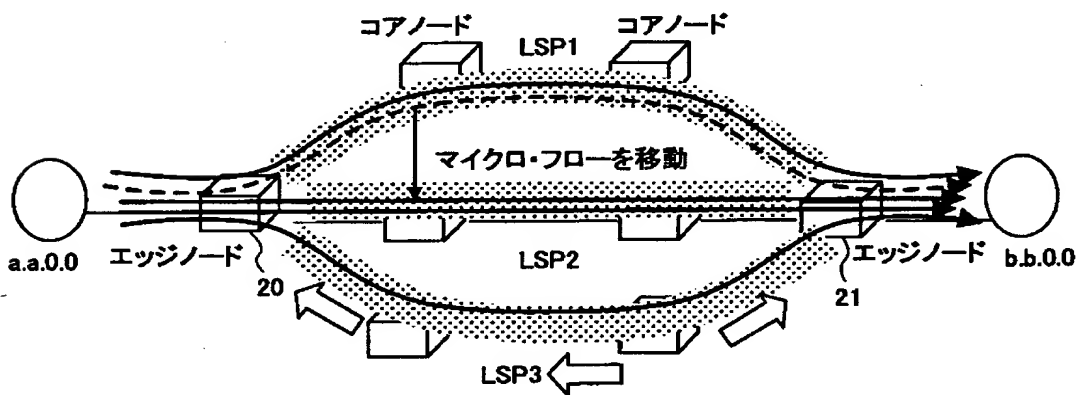
【図 3】

従来のトラフィックエンジニアリング方式を説明するための図



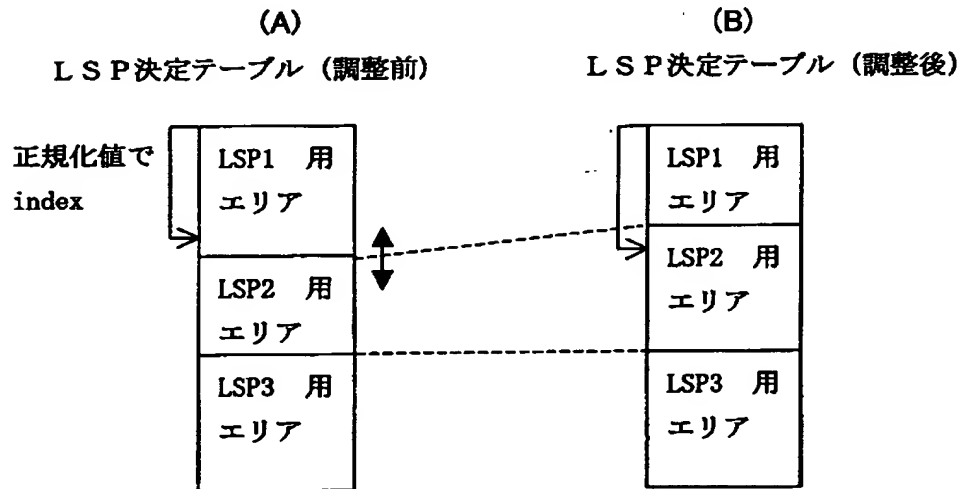
【図 4】

従来のトラフィック分散方式を説明するための図



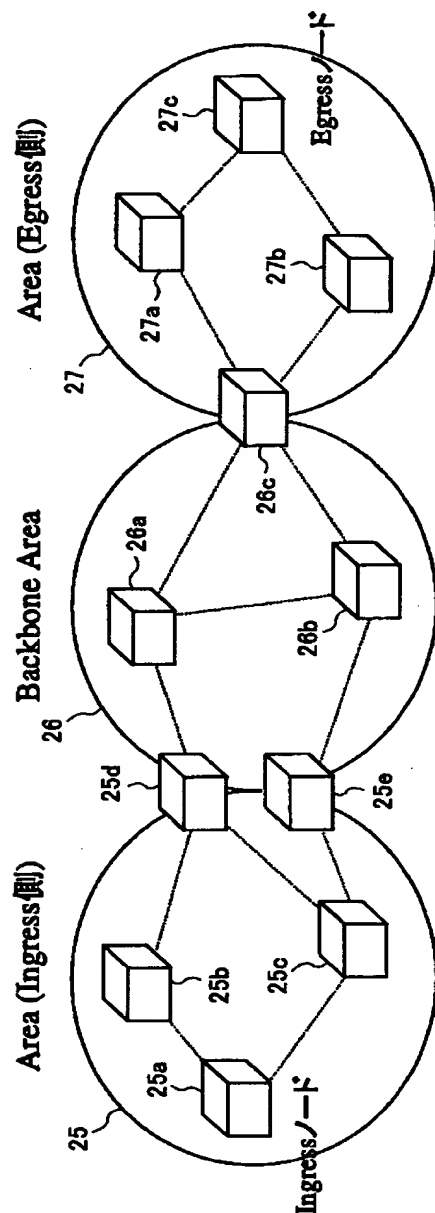
【図 5】

L S P 決定テーブルの L S P の境界の移動を説明するための図



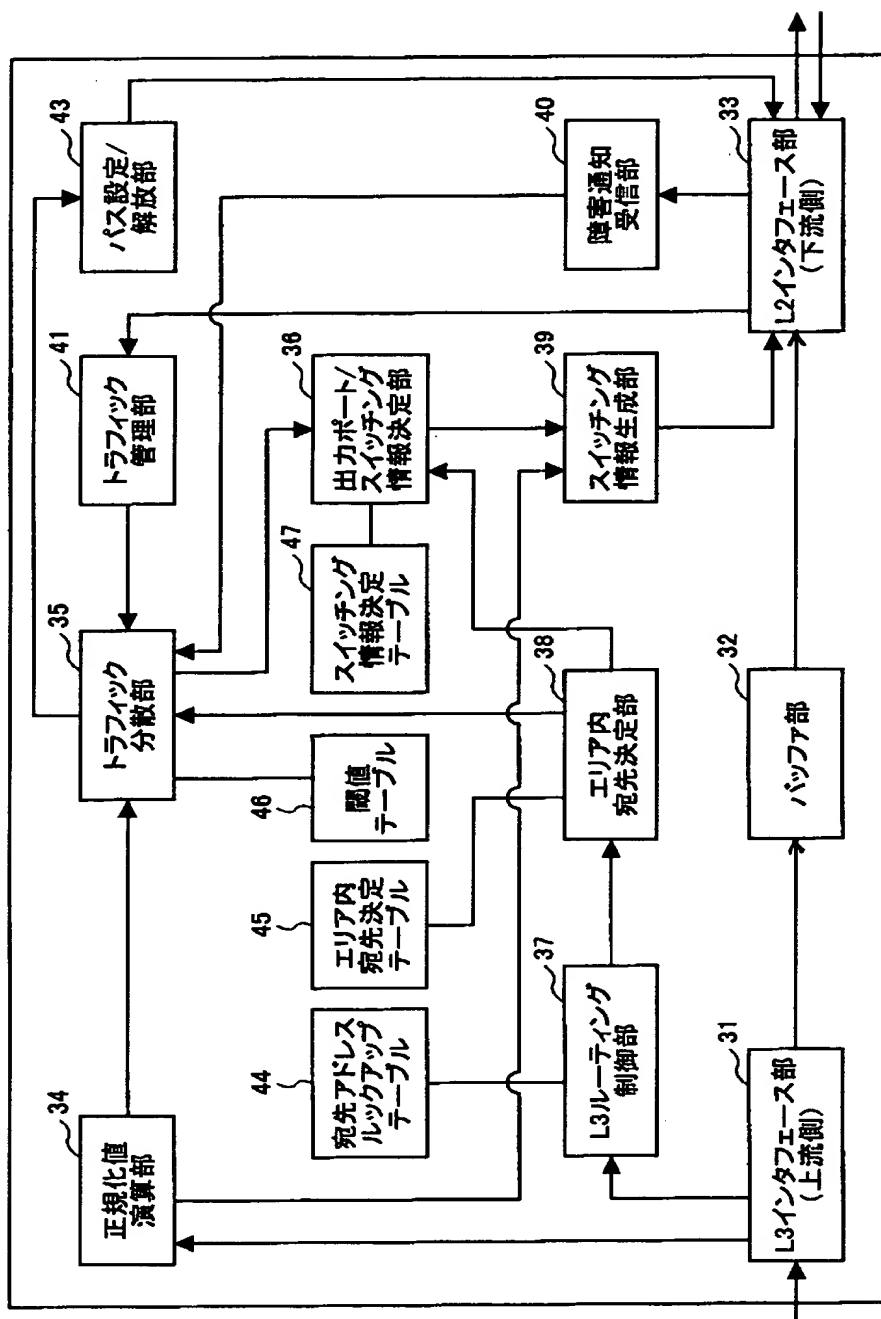
【図 6】

本発明のトラフィックエンジニアリング方法を適用した
ネットワークシステムの一実施例の構成図



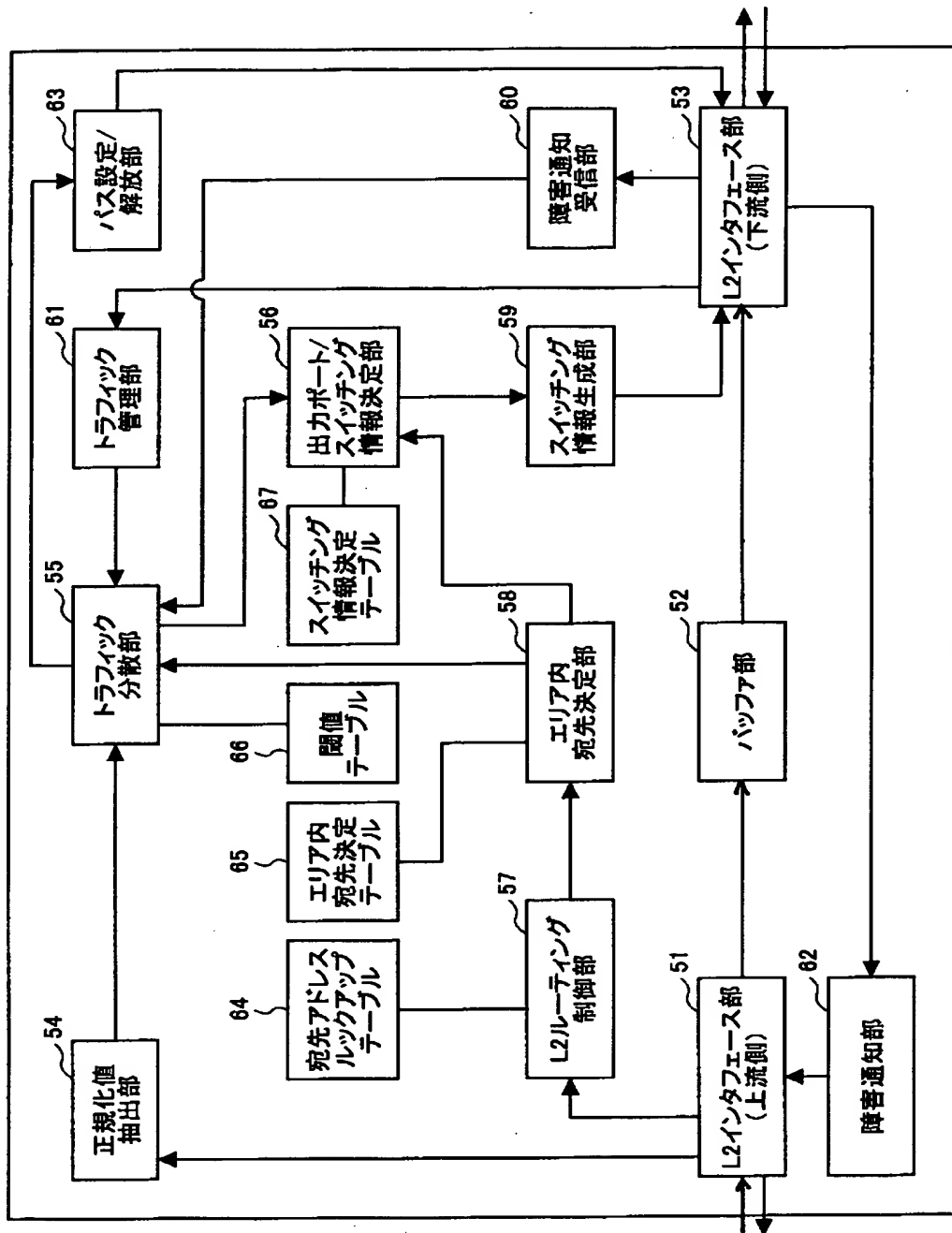
【図 7】

本発明方法に適用されるエッジノードの一実施例のブロック構成図



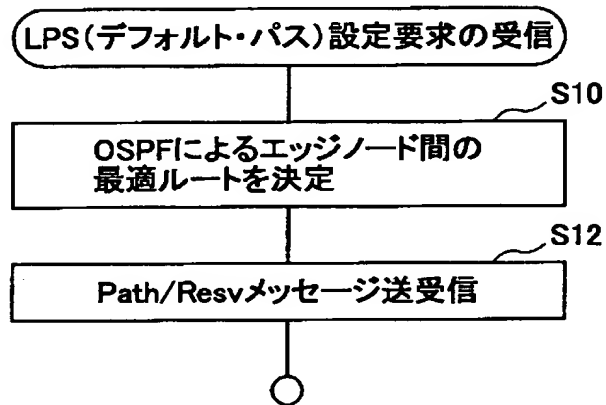
【図 8】

本発明方法に適用されるコアノードの一実施例のブロック構成図



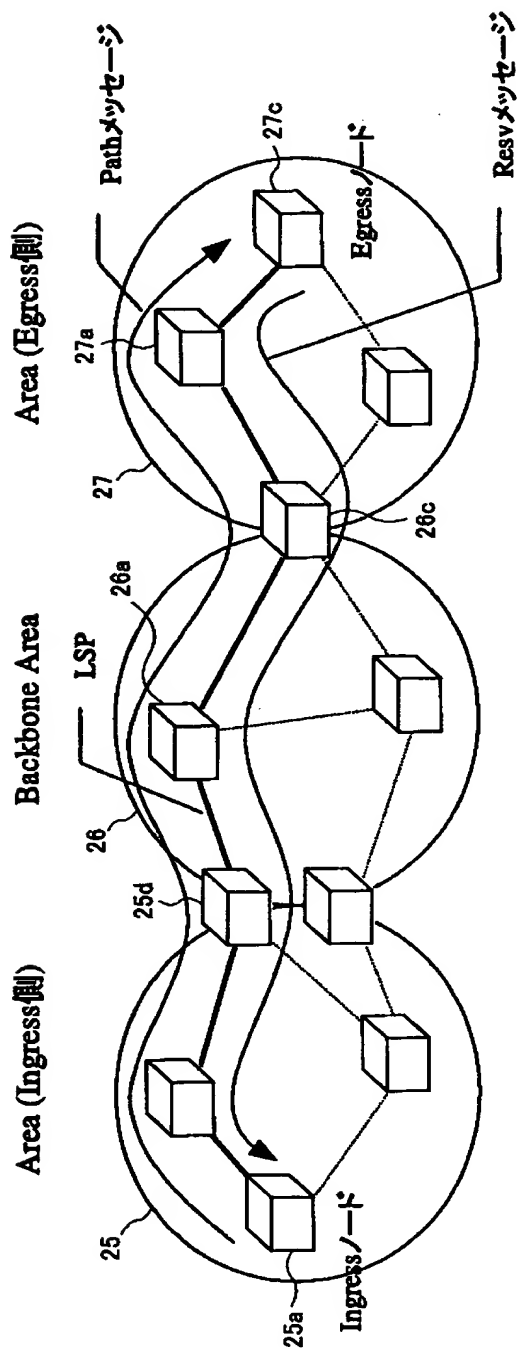
【図 9】

デフォルトパス設定処理の一実施例のフローチャート



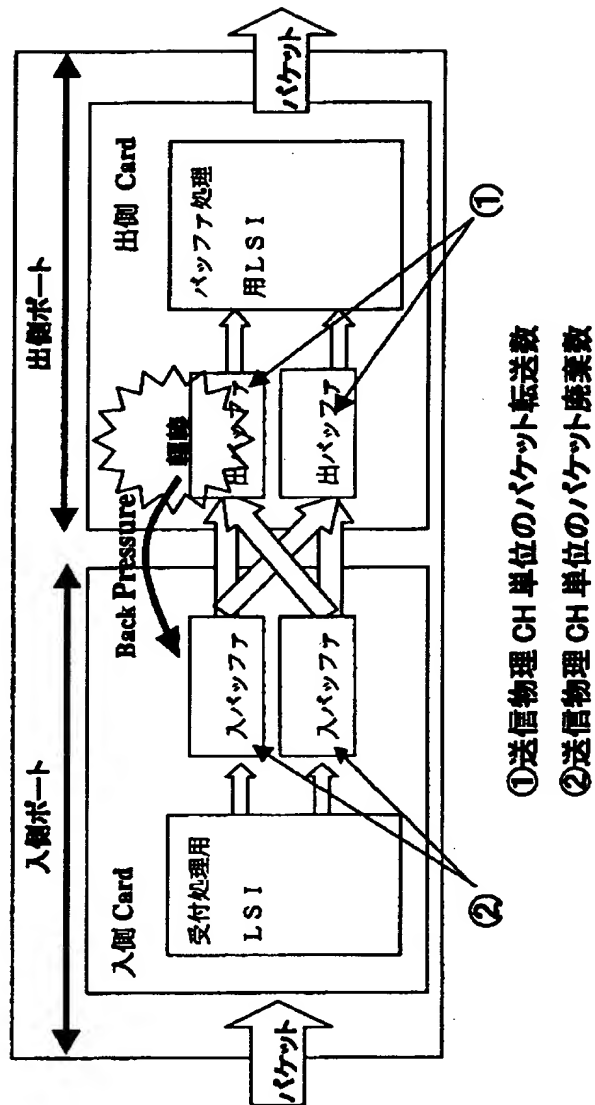
【図10】

デフォルトパス設定を説明するための図



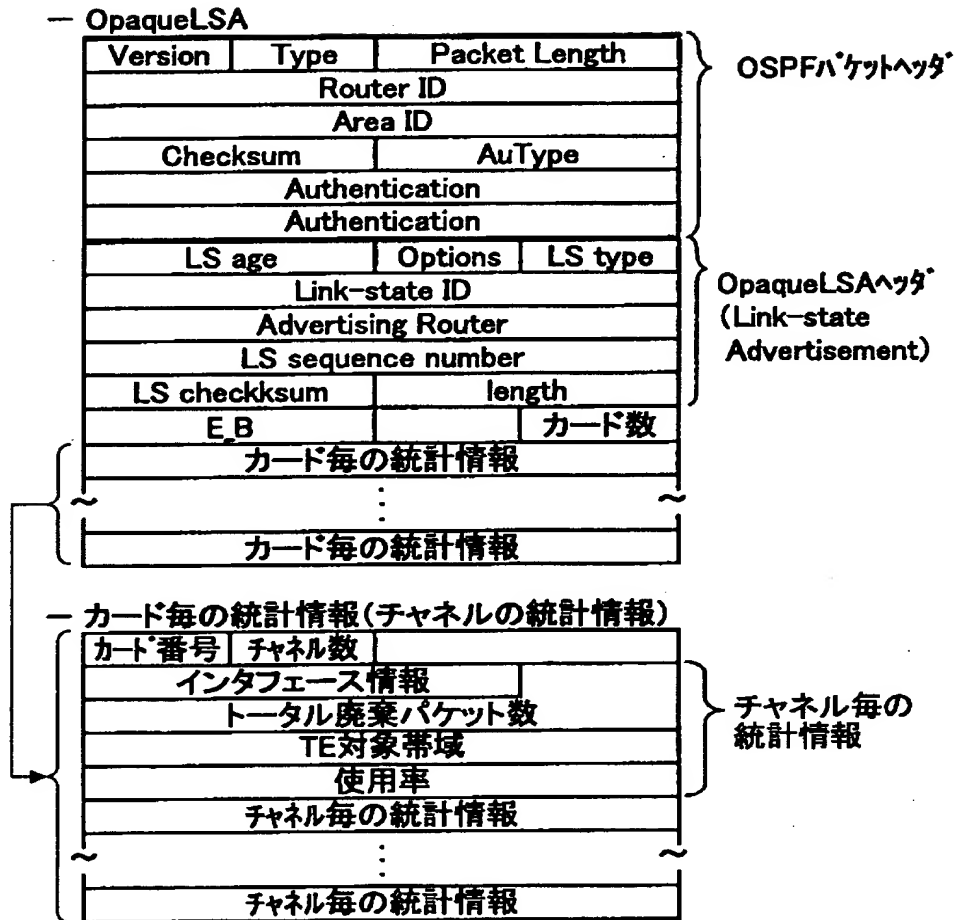
【図 11】

ハードウェアによるトラフィック情報収集を示す図



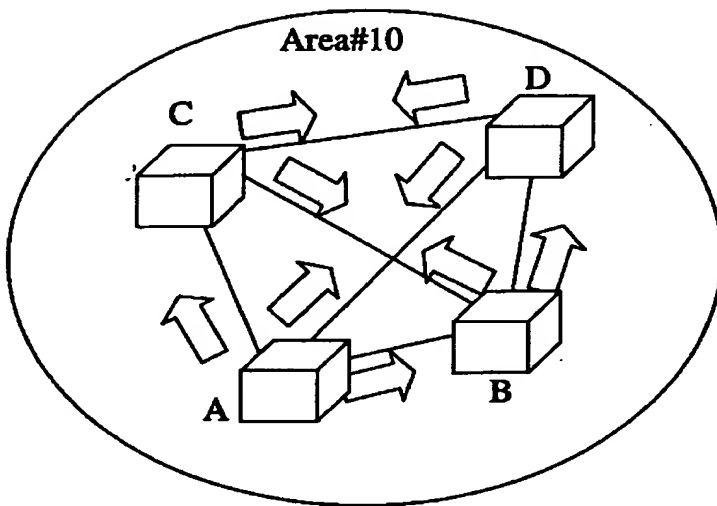
【図 12】

OSPFのOpaqueLSAのフォーマットを示す図



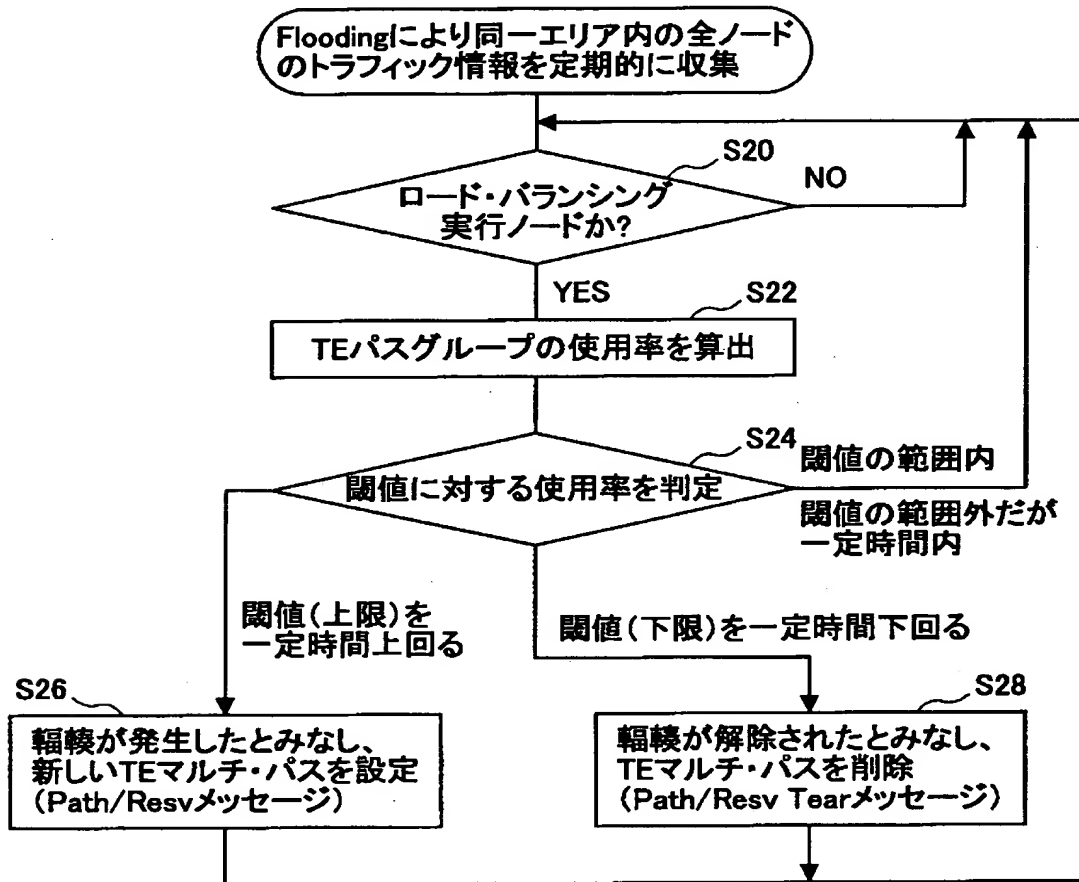
【図 1 3】

広告の様子を示す図



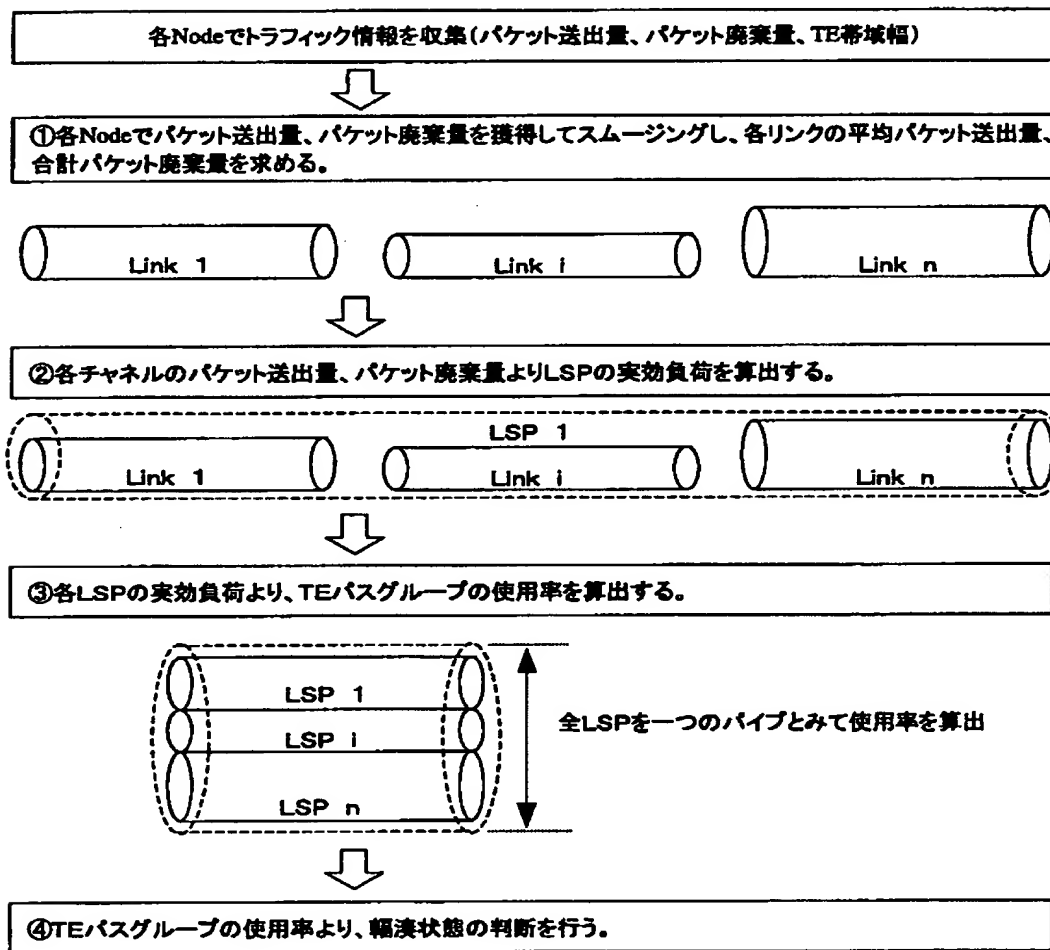
【図 1 4】

ロードバランシング処理の 1 実施例のフローチャート



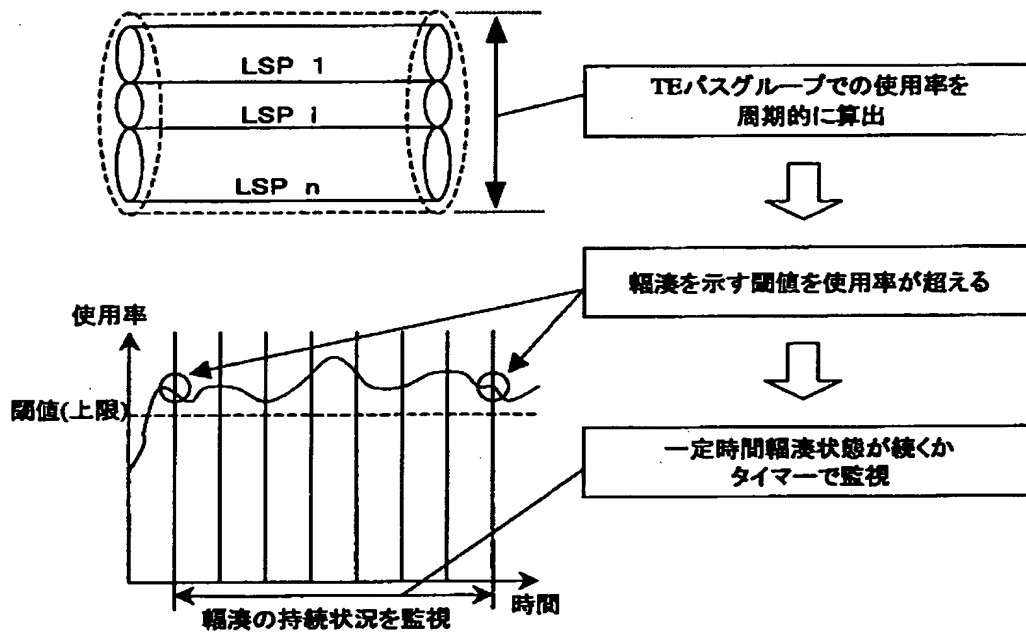
【図 1 5】

TEパスグループの使用率算出を説明するための図



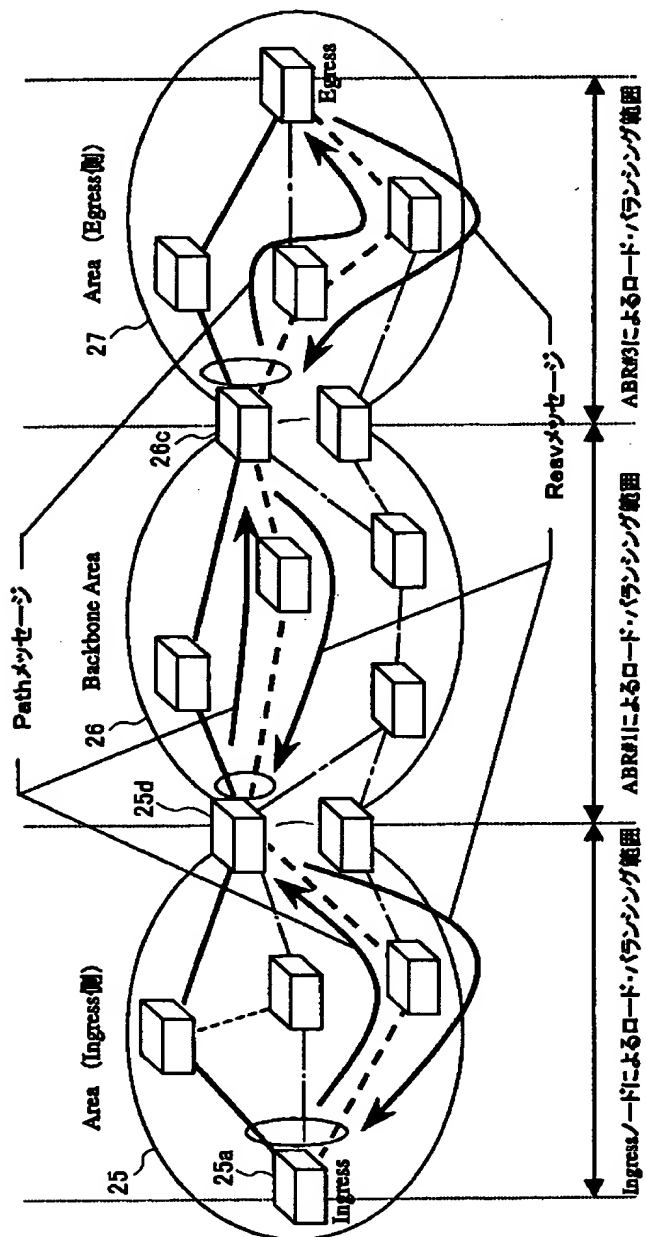
【図 16】

輻輳状況の監視と判定を説明するための図



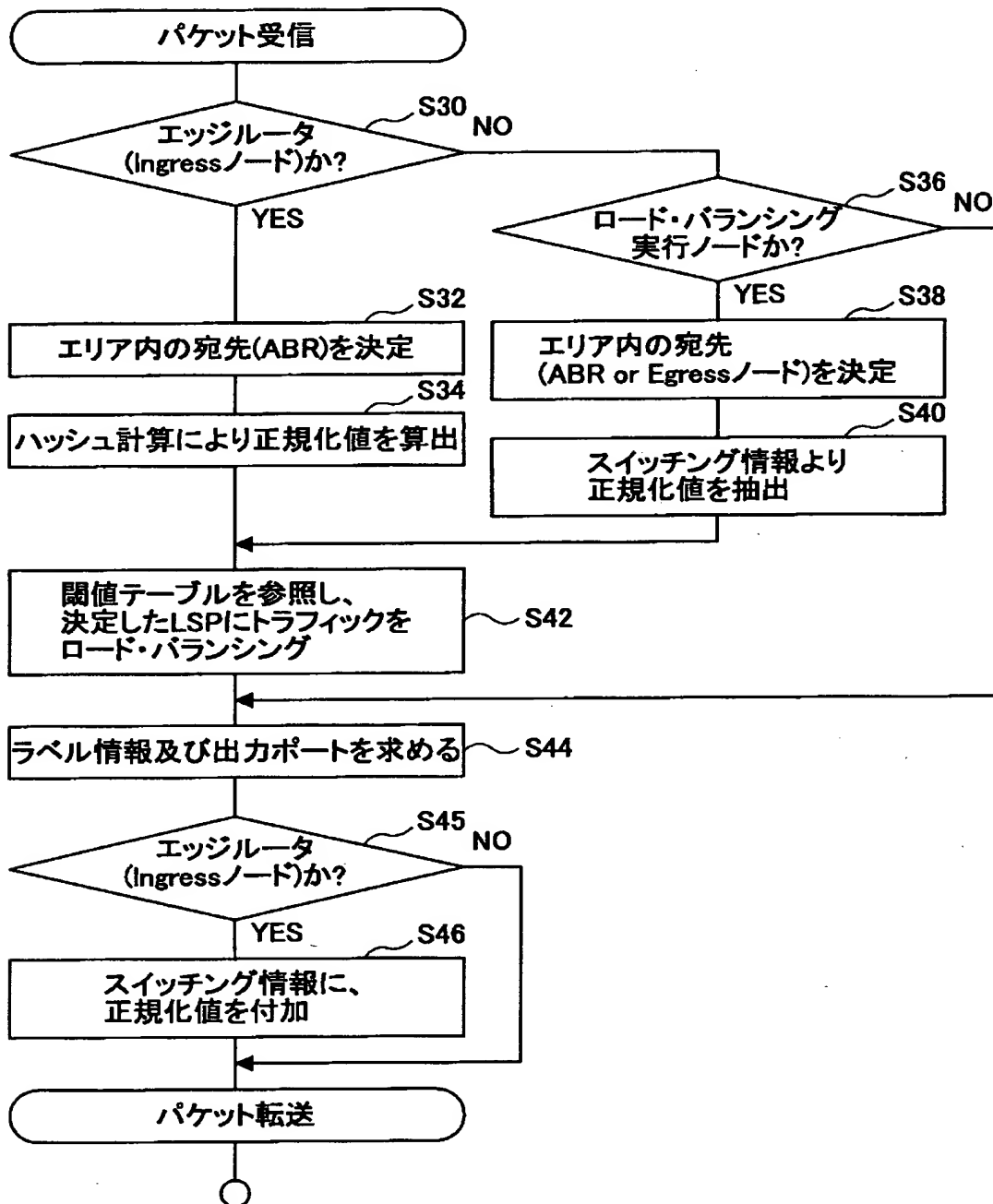
【図 17】

エリア単位のTEマルチパスの設定を示す図



【図 18】

ロードバランシング処理の第2実施例のフローチャート



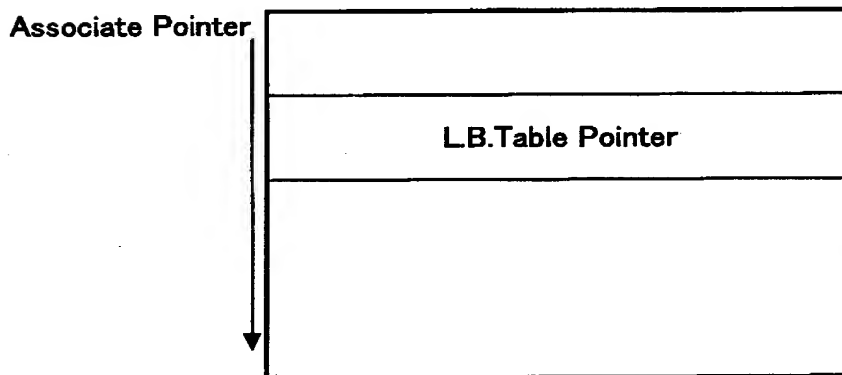
【図 1 9】

宛先アドレスルックアップテーブル 4 4 の構成を示す図

IP Source Address	IP Destination Address	Associate Pointer

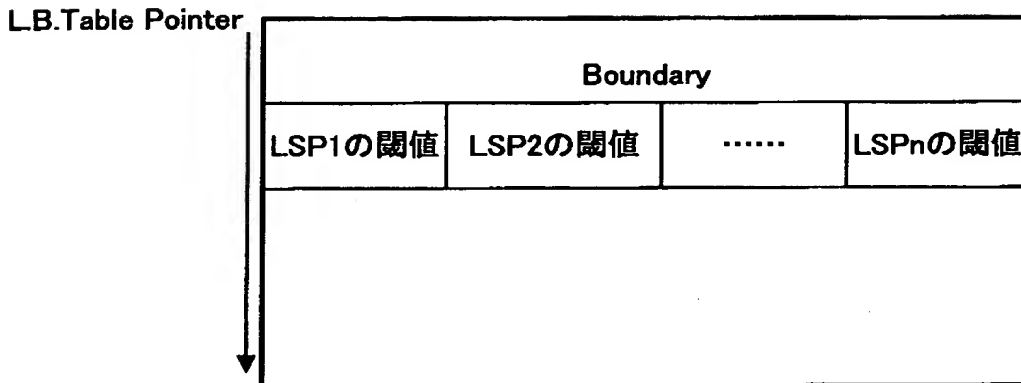
【図 2 0】

エリア内宛先決定テーブル 4 5 の構造を示す図



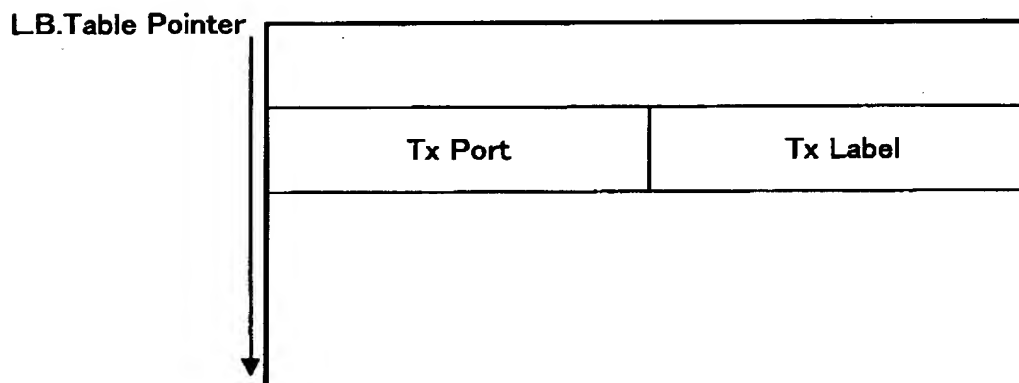
【図 2 1】

構造の閾値テーブル 4 6, 6 6 の構造を示す図



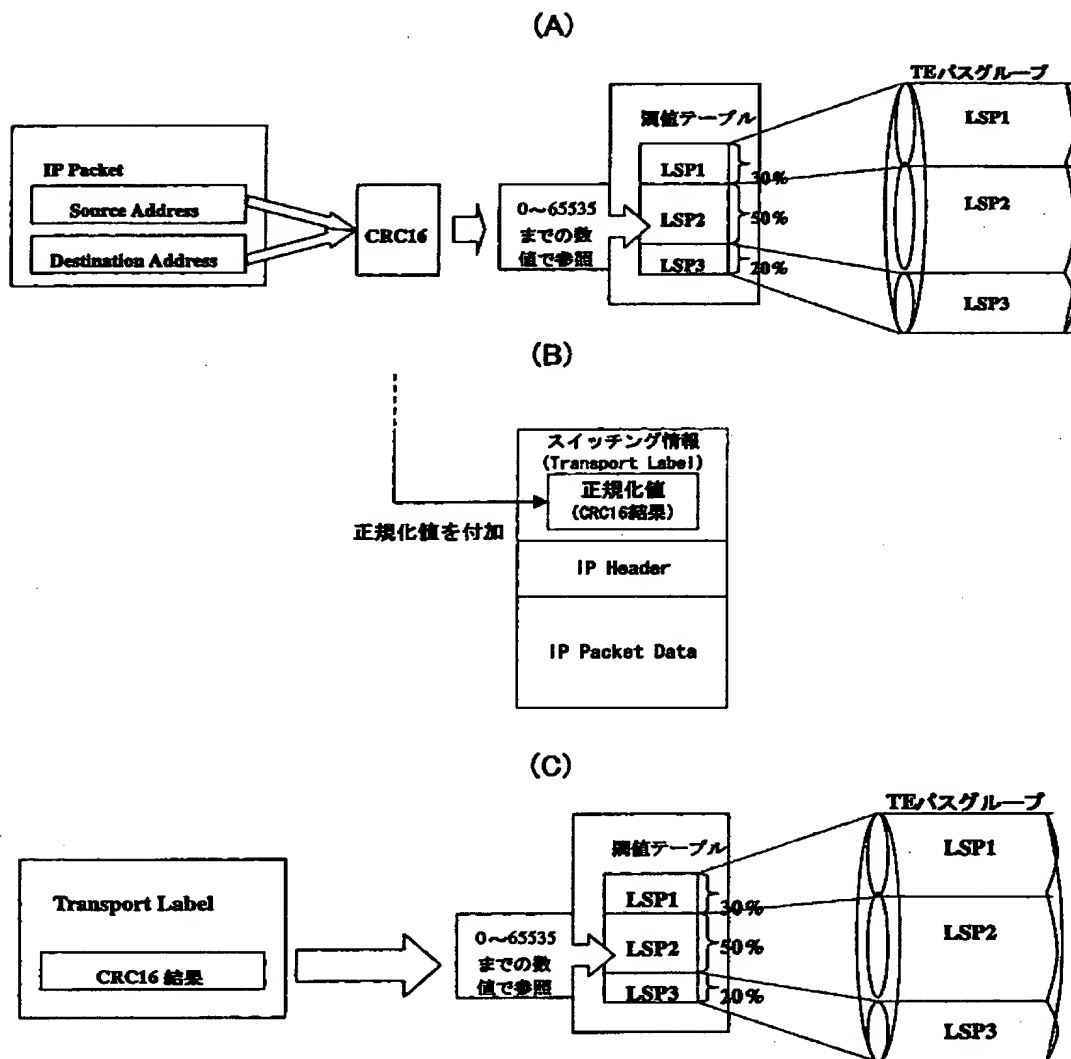
【図 2 2】

スイッチング情報決定テーブル 4 7, 6 7 の構造を示す図



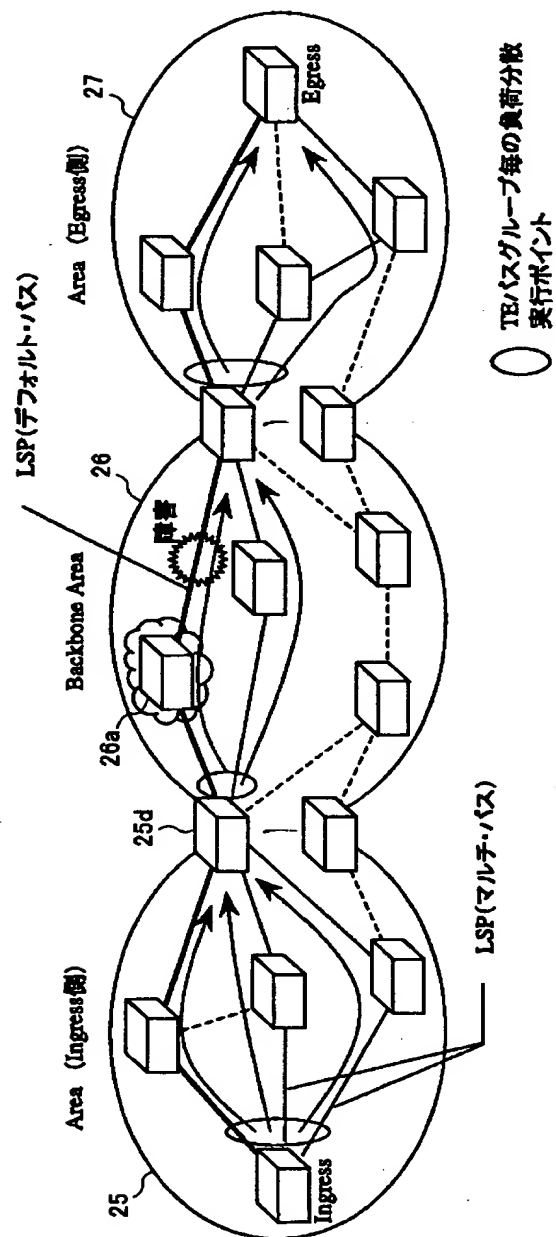
【図 23】

付加分散実施を説明するための図



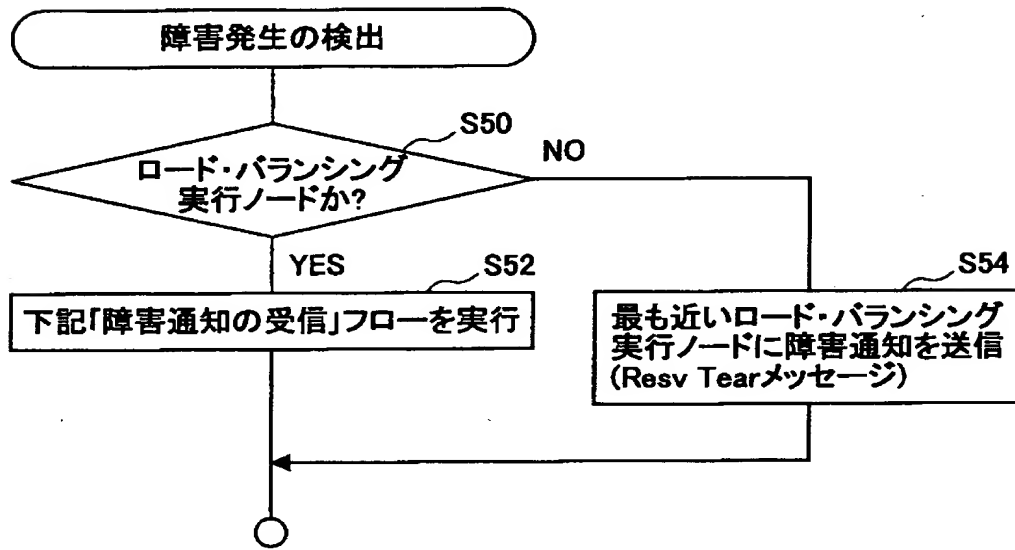
【図 24】

ロードバランシング中での障害検出を説明するための図



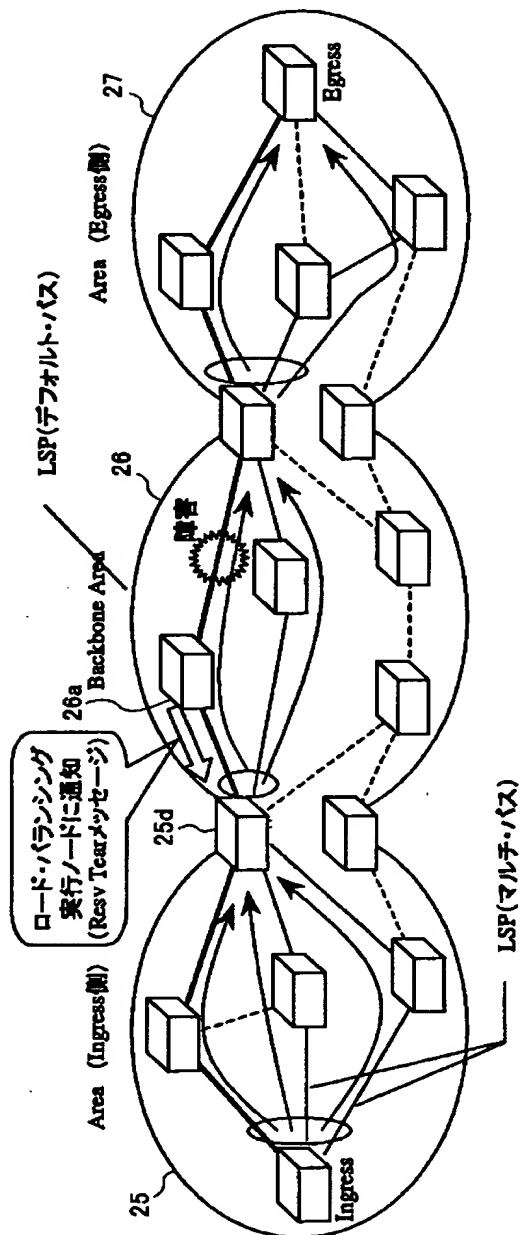
【図 2 5】

障害を検出したノードの障害通知部 6 2 が実行する
障害検出処理の一実施例のフローチャート



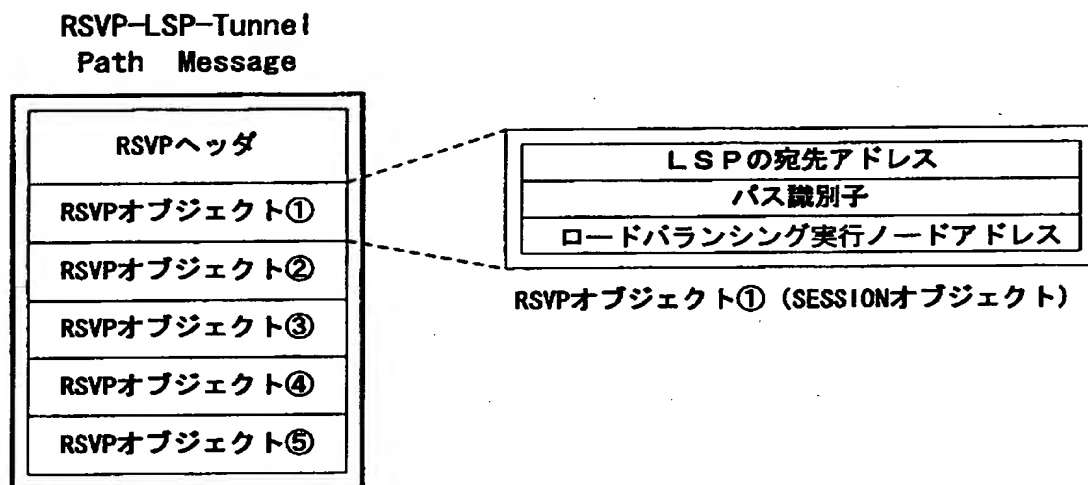
【図 26】

ロードバランシング実行ノードへの障害通知を説明するための図



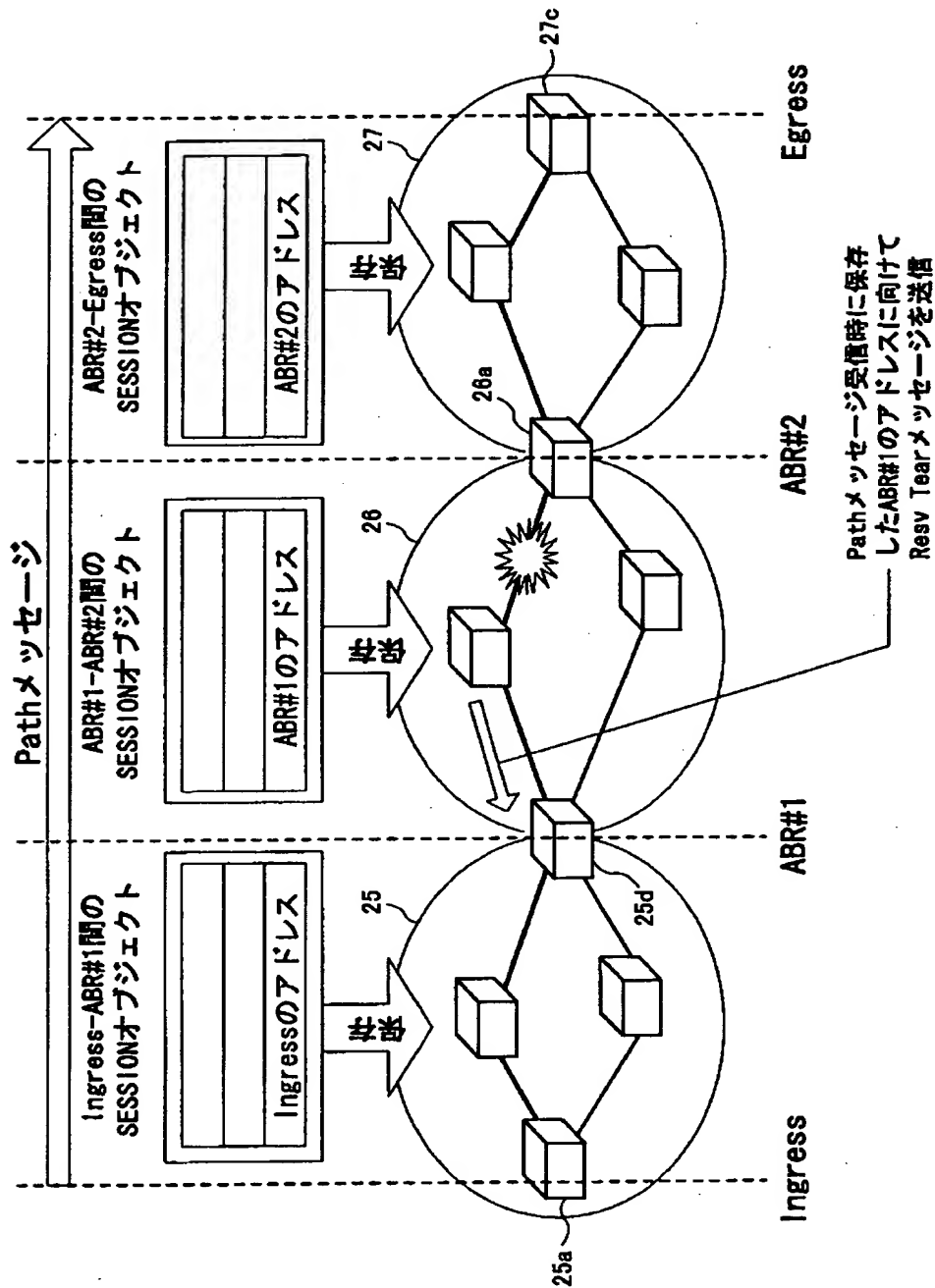
【図 27】

RSVP-LSP-Tunnelの
Resv Tearメッセージの構造を示す図



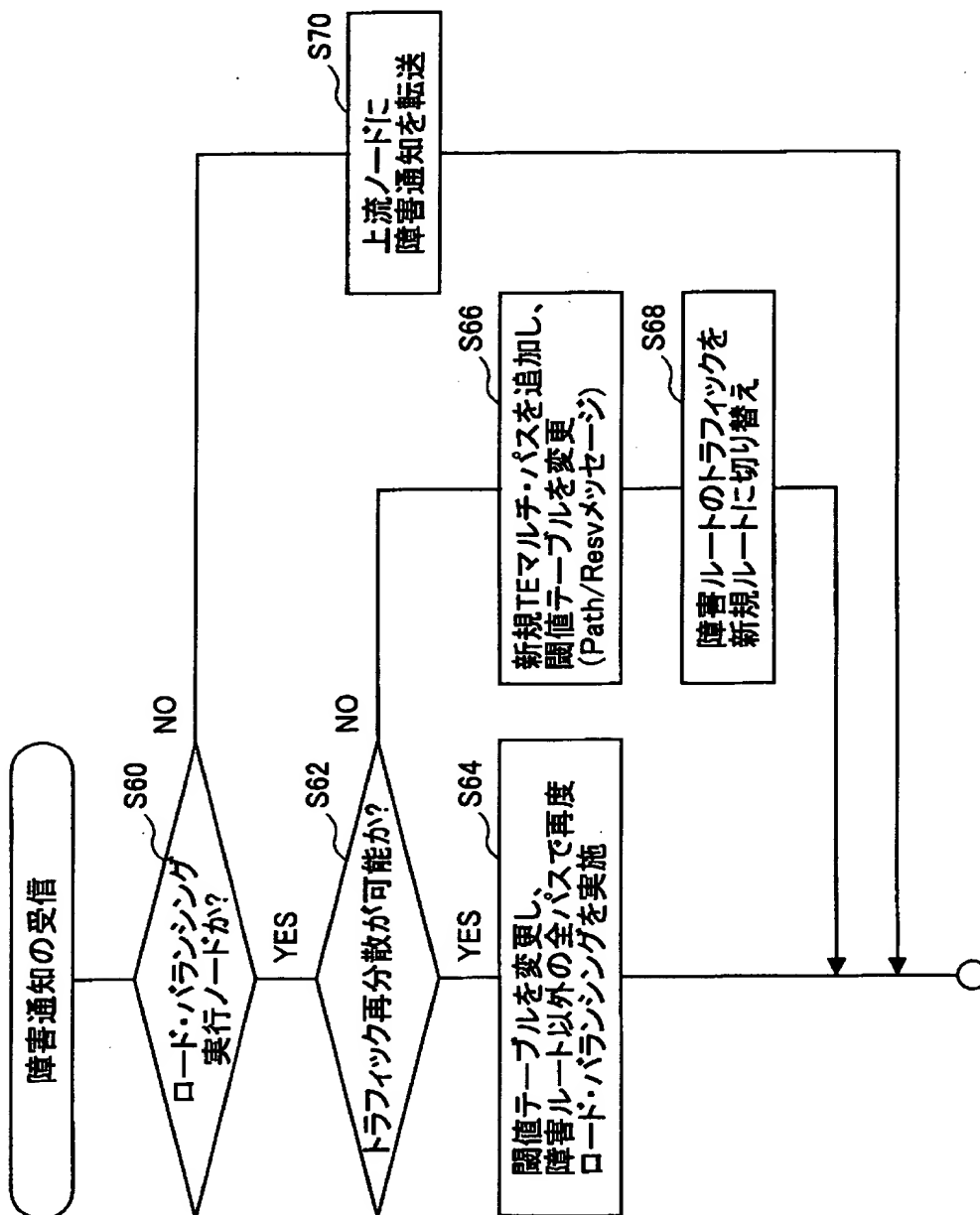
【図 28】

各ノードへの障害通知を説明するための図



【図 29】

障害通知を受信したロードバランシング実行ノードが実行する
障害通知受信処理の一実施例のフローチャート

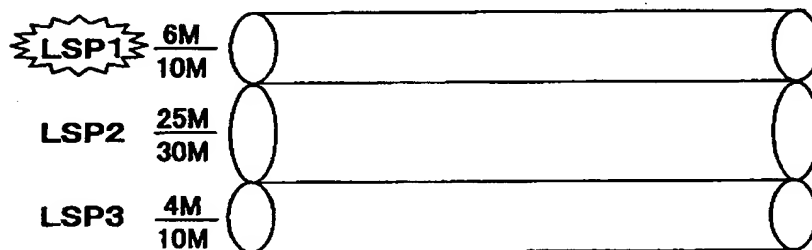


【図 30】

再分散によるトラフィックロス発生判定を説明するための図

(A)

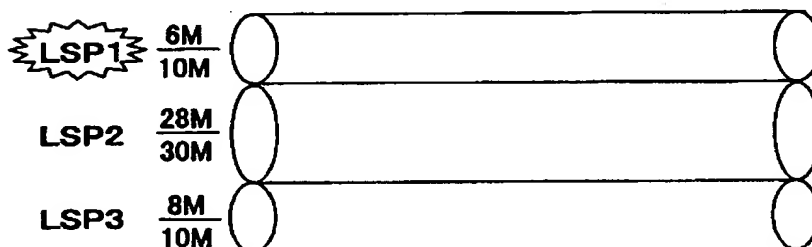
(パターン1: 該障害以外のルートへの再分散が可能な場合)



全体(LSP1/2/3)の空き帯域 - LSP1の空き帯域
 $= 15M - 4M = 11M$
 が、LSP1の実行負荷(6M)よりも大きいので
 トラフィック・ロスが発生せずに再分散可能と判定

(B)

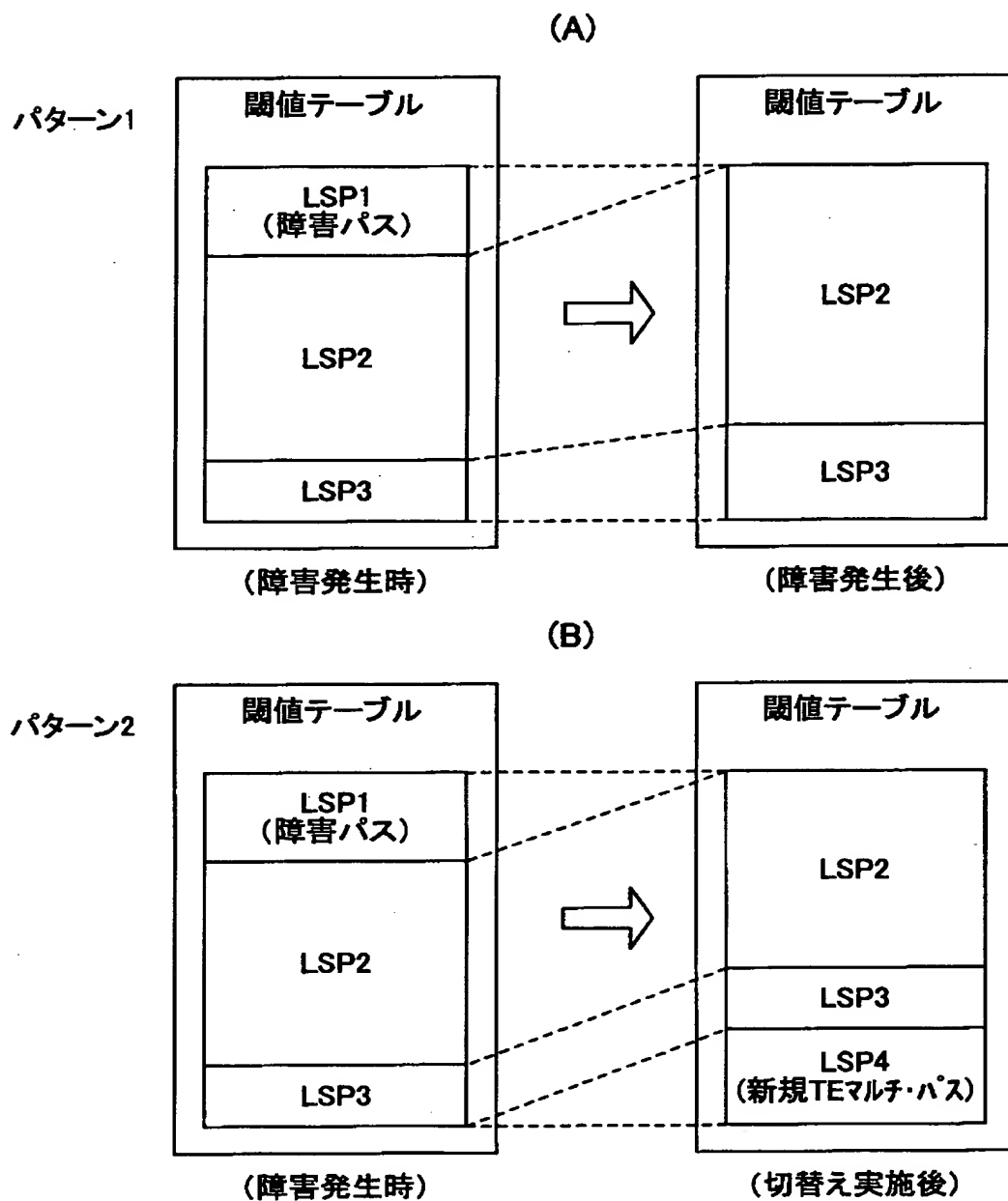
(パターン2: 該障害以外のルートへの再分散不可能な場合)



全体(LSP1/2/3)の空き帯域 - LSP1の空き帯域
 $= 8M - 4M = 4M$
 が、LSP1の実行負荷(6M)よりも小さいので
 トラフィック・ロスが発生せずに再分散不可と判定

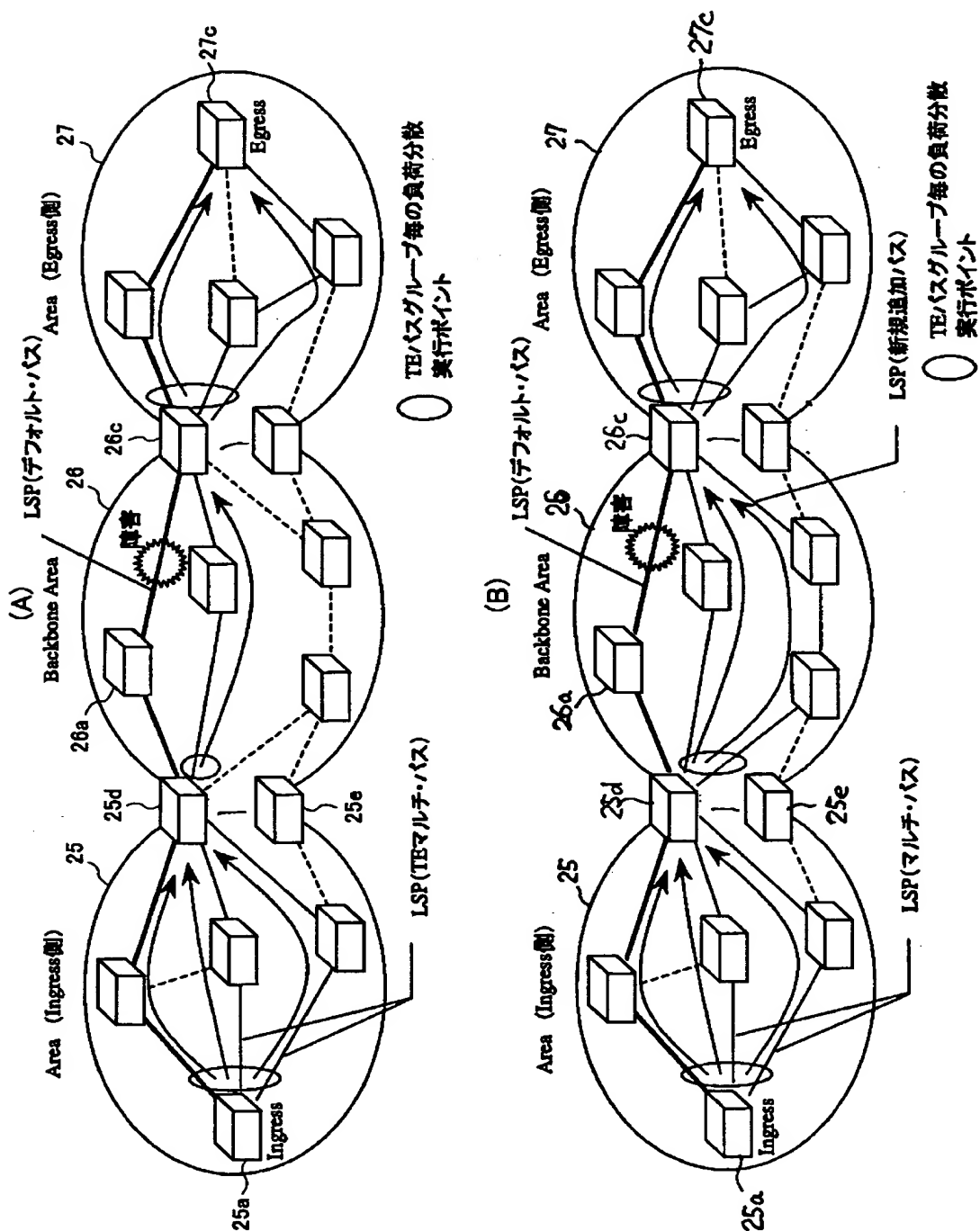
【図 3 1】

再分散による閾値テーブルの変化を説明するための図



【図 3 2】

障害検出後の負荷分散を説明するための図



【書類名】 要約書

【要約】

【課題】 本発明は、大規模ネットワークであっても高速にロードバランシングを実行し、障害が発生した場合も障害ルートのトラフィックロスが高速に救済することができるトラフィックエンジニアリング方法及びそれを用いたルータ装置を提供することを目的とする。

【解決手段】 ネットワーク全体が複数ノードの集合体であるエリアに分割されており、ネットワーク全体の資源の最適化を行うトラフィックエンジニアリング方法において、各エリア毎にエリア内で閉じたロードバランシングを実行するため、大規模ネットワークであってもロードバランシング実行ノードが必要とするメモリ容量を大幅に削減でき高速にロードバランシングを実行することができる。

【選択図】 図4

出 願 人 履 歴 情 報

識別番号 [000005223]

1. 変更年月日 1996年 3月26日

[変更理由] 住所変更

住 所 神奈川県川崎市中原区上小田中4丁目1番1号
氏 名 富士通株式会社